Precise measurements of time delays in gravitationally lensed quasars for competitive and independent determination of the Hubble constant

THÈSE N^{O} 8038 (2017)

PRÉSENTÉE LE 8 DECEMBRE 2017
À LA FACULTÉ DES SCIENCES DE BASE
LABORATOIRE D'ASTROPHYSIQUE
PROGRAMME DOCTORAL EN PHYSIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Vivien François BONVIN

acceptée sur proposition du jury:

Prof. H. M. Rønnow, président du jury
Prof. G. Meylan, Dr F. Courbin, directeurs de thèse
Prof. F. Bouchet, rapporteur
Prof. J. Wambsganss, rapporteur
Dr S. Sibiryakov , rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2017

Remerciements

Il est de bon aloi de remercier les gens m'ayant accompagné tout au long de ma thèse. La plupart étant francophones ou maîtrisant le français, je me permets donc d'écrire ici dans la langue de Molière. Je commencerai donc par remercier le fraîchement titré Prof. Frédéric Courbin, qui m'a supervisé ces quatre années durant. Malgré un agenda plus que chargé, Fred à toujours su ménager du temps pour moi et mon travail, me donnant (le plus souvent) de très bon conseils me permettant ainsi de progresser efficacement et avec enthousiasme. Je tiens aussi à remercier Malte Tewes qui, tant au travers de nos discussions que du travail qu'il a accompli pour COSMOGRAIL, m'a beaucoup apporté. Enfin, je remercie le Prof. Georges Meylan qui m'a accordé sa confiance en m'offrant ce poste de thèse au LASTRO.

Au-delà des livres, des cours et des séminaires en auditoire, l'apprentissage se fait bien souvent au travers d'essais et d'erreurs. La discussion y joue un rôle prépondérant, permettant de confronter ses idées au jugement de ses pairs. N'ayant pas forçément envie de me mettre mes superviseurs de thèse à dos à coup de questions parfois autant biscornues qu'asbconses, j'ai trouvé oreille attentive auprès de mes confrères doctorants et postdocs. Je ne saurai probablement jamais s'ils m'écoutèrent vraiment avec intérêt ou s'ils eussent préféré que je me tusse mais quoi qu'il en soit, j'ai grandement apprécié de pouvoir échanger avec Thibaut Kuntzer, Markus Rexroth, Olga Tihhonova, Danka Paraficz, Aurélien Wyttenbach, Sébastien Peretti, Rémy Joseph et tant d'autres.

For those who do not speak French but might still ctrl+F their name: I had the chance to collaborate with a fine bunch of experts in time-delay cosmography. It took me a lot of efforts to catch-up with them and try to match their standards, but I definitely grew a lot in the process and I am very thankful for that. Praise to Sherry Suyu, Phil Marshall, Tommaso Treu, Dominique Sluse, Kenneth Wong, Thomas Collett and all the members of the H0LiCOW collaboration. Doing that time-delay cosmography thing with you was a real thrill, and something tells me that an even more exciting future lies ahead.

Etre bien entouré au-delà de son lieu de travail est d'une importance capitale. J'ai la chance de partager ma vie avec Ewelina Obrzud, une fille aux qualités innombrables, notamment celle de me remettre les pieds sur terre de temps à autre. Je suis tout aussi chanceux d'avoir bénéficié d'un soutien sans faille depuis toujours de mes parents Marianne et Raymond Bonvin, pour qui je suis sans aucun doute le plus grand fournisseur de métaphores culinaires et jardinières pour expliquer comment fonctionne l'Univers.

Je ne peux pas dire avoir été passionné d'astronomie - ou même de physique - depuis ma plus tendre enfance. Mon truc à moi, alors adolescent, c'était plutôt les jeux vidéos et le tennis. Pourtant, on ne se retrouve pas à concourir pour le titre de docteur en Astrophysique par hasard; l'influence de mon entourage y était forcément pour quelque chose. Je pense tout d'abord à ma grande soeur, Camille Bonvin: c'est probablement grâce à elle que je me suis retrouvé à étudier la physique à l'EPFL ("La physique c'est fantastique, tu verras"). Et puis, pour tenir cinq ans d'études parfois assez brutales il m'a fallu faire preuve d'une bonne dose d'abnégation. Je ne l'aurais probablement pas eue si je n'avais développé ma résistance à l'effort en suivant Michel Siegenthaler en haute montagne. Marcher des heures durant dans le froid et dans la nuit pour contempler durant quelques instants, au-dessus des nuages, la beauté d'un lever de soleil, a de quoi marquer plus d'un adolescent. Enfin, mon premier contact avec l'astronomie s'est faite sur le tard, à l'observatoire François-Xavier Bagnoud à St-Luc lors d'une semaine d'initiation organisée par Jean-Claude Pont. Les bons moments passés durant trois étés consécutifs m'ont ouvert à la plus vieille des sciences, et j'ose affirmer aujourd'hui que ce fut pour le meilleur.

La physique c'est fantastique, en effet. Appréciant la compétition (mais pas la défaite), je me suis retrouvé en 2011 à participer à l'International Physicists' Tournament (IPT), une compétition pour étudiants physiciens basés sur la présentation de problèmes de recherche résolus au préalable. Ce qui ressemble à s'y méprendre a une bête résolution d'exercice va en fait bien plus loin; c'est seulement après avoir tenté de convaincre un auditoire suspicieux de la possibilité d'allumer une bougie avec un arc et des flèches[1] ou de s'orienter dans l'espace en utilisant uniquement une étoile à neutron[2] que j'ai vraiment compris l'intérêt d'une telle compétition. En poussant les participants à effectuer leur propre recherche et à en tirer leur propres conclusions, l'IPT offre une première expérience dans le monde de la recherche scientifique. Y prendre part a certainement influencé mon choix de me lancer un doctorat, mais le meilleur était encore à venir. De fil en aiguille, je me suis retrouvé à organiser le tournoi en Suisse sur le site de l'EPFL, et en suis devenu depuis 2014 le Président. Lors de ces six dernières années, le tournoi a grandi pour passer d'une compétition européenne à un tournoi vraiment international, réunissant 15 pays différents provenant de 4 continents. Au-delà du plaisir de voir grandir cet événement d'année en année, en être en charge m'a permis de saisir à quel point un groupe d'individus motivés peuvent changer les choses autour d'eux. Etre capable de réunir des étudiants Chinois, Américains, Russes et Vénézuéliens en Suède discutant de la manière la plus optimale de construire une tour en gélatine, simplement pour le plaisir du débat est quelque chose de très particulier. Cette expérience extraordinaire a très certainement fortement influencé ma manière de concevoir mon travail de doctorat. Quoi

---

[1] En réalité, c'est impossible.

[2] Idem.

qu'il en soit, L'IPT et moi-même ne serions pas ce que nous sommes sans l'engagement dans cette aventure de Maxime Harazi, Daniel Suchet, Lucy Voronina, Vladimir Vanovskiy, Andreas Isacsson et j'en oublie. Merci pour ces moments passionnants.

Si mon travail consitait à casser des cailloux, j'aurais pu le faire sans état d'âme. Mais il m'a fallu réfléchir, du moins de temps à autre. Et dans de tels cas, il ne m'était simplement pas possible de dissocier les joies et les peines de la vie de tous les jours de mon activité cérébrale. La qualité de ma thèse reflète ainsi ma qualité de vie; à tout ceux qui m'ont cotoyé ces quatre années durant, il y a un peu de chacun de vous dans cette thèse. Et si au final mes pairs me jugent digne de recevoir le titre de Docteur, c'est en partie grâce à vous. Merci à tous.

Vivien

Abstract

During these last decades, by virtue of observations, the *Standard Cosmological Model* has emerged, providing a description of the Universe's evolution using a minimal set of independent constraints - the *cosmological parameters*. Among them is the expansion rate of the Universe, the so-called Hubble constant or H_0, first measured by Lemaître in 1927. The century that followed this cornerstone measurement saw numerous attempts to refine the initial value, and for good reason: a precise and independent measurement of H_0 will bring strong constraints on the cosmological models. It could notably help the astronomers to better understand the nature of dark energy, thus making it one of the most sought-after prizes in modern cosmology.

My work at the Laboratory of Astrophysics of EPFL is embedded in this context. I am part of the COSMOGRAIL and H0LiCOW[3] collaborations, aiming to measure the Hubble constant with the highest level of precision using time-delay cosmography, a method based on the theory of strong gravitational lensing. This effect occurs when an observer looks at a light source located behind a massive foreground galaxy. The mass of the galaxy acts similarly to an optical lens and focuses the light rays emitted by the source. As a consequence, multiple *lensed* images of the source appear around the *lens* galaxy. If the luminosity of the source changes over time, the variations will be seen in all the lensed images but with a temporal delay due to the different travel paths of the light rays. By carefully monitoring the luminosity variations of each lensed image, one can precisely measure the temporal delays between them. Combined to high-resolution observations of the foreground galaxy and its surroundings, it is possible to directly measure the Hubble constant upon the sole assumption that the General Relativity is correct.

Since more than 13 years, COSMOGRAIL monitors dozens of lensed quasars to produce high-quality light curves and time-delay measurements. During these last four years, I took care of the monitoring schedule, continuous data reduction and time-delay measurements through the development of curve-shifting techniques. I produced light curves and measured time delays on a variety of lenses. After more than a decade of endeavours, COSMOGRAIL and H0LiCOW finally revealed their measurement of the expansion rate of the Universe from a blind analysis of three lensed sources. I had the privilege to be the lead author of the publication presenting our measurement of the Hubble constant, $H_0 = 71.9^{+2.4}_{-3.0}\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$, at 3.8% precision in the Standard Cosmological Model. Such a precision allows a direct comparison

---

[3] see www.cosmograil.org and www.h0licow.org

with the results of the *distance ladder* technique in the local Universe and the *Planck* satellite Cosmic Microwave Background observations in the distant Universe, both of which being currently in a significant tension of unknown source.

Key words: Cosmological parameters; Time-delay cosmography; Strong gravitational lensing

Résumé

Au cours des dernières décennies et en vertu de nombreuses observations, le *Modèle Cosmologique Standard* a émergé, fournissant une description de l'évolution de l'Univers au cours du temps et ce en utilisant un ensemble minimal de contraintes que sont les *paramètres cosmologiques*. Parmis eux se trouve le taux d'expansion de l'Univers, aussi nommée constante de Hubble - ou H_0 - mesuré pour la première fois par Lemaître en 1927. Le siècle suivant cette observation pionnière a vu défiler de nombreuses tentatives d'améliorer l'estimation initiale de Lemaître, et pour cause: une mesure à la fois précise et indépendante de H_0 permet de contraindre efficacement les modèles cosmologiques actuels. Notamment, une telle mesure aiderait les astronomes à mieux comprendre la nature de l'énergie sombre, faisant ainsi de la constante de Hubble l'un des paramètres les plus prisés de la cosmologie moderne.

Mon travail au Laboratoire d'Astrophysique de l'EPFL s'inscrit dans ce contexte. Je fais partie des collaborations COSMOGRAIL et H0LiCOW[4] qui cherchent à mesurer la constante de Hubble avec un haut niveau de précision en utilisant une méthode basée sur les délais temporels induits par l'effet de lentille gravitationnel fort. Lorsqu'un observateur regarde une source de lumière située au-delà d'une galaxie massive d'avant-plan, la masse de galaxie agit de manière similaire à une lentille optique et dévie les rayons lumineux émis par la source. En conséquence, plusieurs images *mirages* de la source apparaissent autour de la galaxie *lentille*. Si la lumière émise par la source varie au cours du temps, ces variations seront alors observées sur chaque image mirage, mais avec un certain décalage temporel dû à la différence de longueur des chemins parcourus par les rayons lumineux. En mesurant régulièrement les variations de luminosité de chaque image mirage, il est possible de mesurer précisément le décalage temporel entre les mêmes variations observées dans les différentes images mirages. Combiné à des observations à haute résolution de la galaxie d'avant plan ainsi que de son entourage, il est alors possible d'en extraire directement la constante de Hubble.

Depuis plus de 13 ans, COSMOGRAIL observe régulièrement des dizaines de quasars lentillés afin d'en tirer des courbes de variation de luminosité des images mirages et d'y mesurer les délais temporels. Au cours des quatre dernières années, je me suis occupé de la planification des observations, de la réduction des données ainsi acquises mais aussi de la mesure de délais temporels qui en découlent. Après plus d'une dizaine d'années d'efforts, COSMOGRAIL et H0LiCOW révélèrent en 2016 leur mesure indépendante de la constante de Hubble, soit $H_0 = 71.9^{+2.4}_{-3.0}\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$, à une précision 3.8%. Une telle précision permet une comparaison

---

[4] voir www.cosmograil.org et www.h0licow.org

directe avec les résultats obtenus via la technique des *échelles de distances* dans l'Univers local ainsi qu'avec les prédictions du satellite *Planck* observant le fond diffus cosmologique, ces deux resultats étant actuellement incompatibles l'un avec l'autre.

Mots-clés: Paramètres cosmologiques; Cosmographie des délais temporels; Effet de lentille gravitationnelle fort

Contents

1 Introduction

"Besieged, the secular response is militant, an atheism volubly affirmed that in its dogmas and its certainties approaches the religious, although armed with nothing more substantial than established scientific fact, itself a changed constituency of shifting ground. The classical and quantum models are persistent in rejecting all attempts at reconciliation, with the string by which they might be bound proving thus far elusive. Insufficiently grasped gravity engenders multiplying entities in its support, exotic states and substances, dark energy, dark matter, necessary beasts arisen from mathematics yet escaping observations. Faith and politics ferment, aided by a fast-propagating yeast of theory and device, and all the architecture of the world's traditions seems erected on an information floodplain, vulnerable to every fresh downpour of data or the bursting banks of ideologies too narrow and slow-moving to accommodate the surge, the inundation of complexity. Despite its evident fatigue, afraid of missing some vital development in this incessant and incendiary pageant, culture dare not close its eyes."

- Alan Moore, Jerusalem -

1.1 General overview

1.1.1 A word of warning

For the sake of clarity, the first part of this introduction will jump back and forth in time, sometimes omitting important parts of the fields discussed only to come back to them later. I have chosen to proceed this way in order to introduce as smoothly as possible the principles required to understand the scope of my work to the reader not familiar with the field. From the birth of General Relativity to the observational proof of the accelerated expansion of the Universe, and including serendipitous discoveries such as the Cosmic Microwave Background, the History of Cosmology is a really deep, sometimes chaotic but in overall fascinating field that could be (and probably have already been) the topic of a whole thesis. I do not have here such ambitions nor abilities. The goal of the introduction that follows in the next twenty pages or so is not to be exhaustive or to delve into details that have, for certainty, their importance in the history of the field. It is rather to present a simplified and pedagogical picture of the context in which I worked to a broader audience. My hopes are that this introduction, that is before all meant to be pedagogical, will catch the attention of a readership larger than the people of the jury compelled to go trough the present document.

1.1.2 The geometry of space and time

My work uses what we call the Strong Gravitational Lensing. This effect is based on the equations of the General Relativity (Einstein, 1915b), that tell us that the space and time we are living in are, in fact, connected. The fabric itself of space and time is altered by the presence of local concentrations of energy, emerging indistinctly from matter or radiation alike. But let us put aside the time part for the moment and focus solely on space. To put the sentence above in simpler words, if a massive body -like a star or a galaxy- is located somewhere in space, then whatever travels besides it will be affected by its presence. We experience that effect ourselves everyday, being attracted at the surface of the Earth, or seeing the Earth orbiting the Sun - it has to be noted that one of the great early successes of Einstein's theory was that it was able to explain with a greater precision the evolution of Mercury's orbit over the years (Einstein, 1915a). But massless objects, like photons, are also affected by the presence of massive objects. Seen with an omniscient perspective, their trajectory when passing close to a massive object are slightly bent towards it. Interestingly, though, if a massless version of you were able to ride these photons it would not notice any change of trajectory, because its perception of space would be affected as well !

To understand it better, let us assume that little folks living in a 2-dimensional plane, like a sheet of paper are experiencing this effect. The massive body located at the center of the sheet will deform it creating what we call a potential well. The deformation occurs in a way that only a being able to experiment a third spatial dimension can perceive, like us right now when looking at the illustration of Fig. 1.1. A trajectory that looks straight when standing on the 2-dimensional sheet is in fact curved when seen from the 3-dimensional space outside. Similarly to the flat people in their tiny houses living on the sheet of paper, we are not able to directly perceive with our human senses the bending of our 3-dimensional space, yet it occurs.

1.1.3 The gravitational lensing effect

We cannot perceive the bending of space nor the dilation and contraction of time when we experience them directly, but we are able to see their effect indirectly. These were predicted by Einstein when he published his theory of General Relativity (Einstein, 1915b, hereafter GR), and were confirmed a few years later by Dyson et al. (1920) with the observations of the position of stars in the sky close in projection to the Sun, during a solar eclipse. When comparing the position of the stars during the eclipse with the ones when the Sun was no more

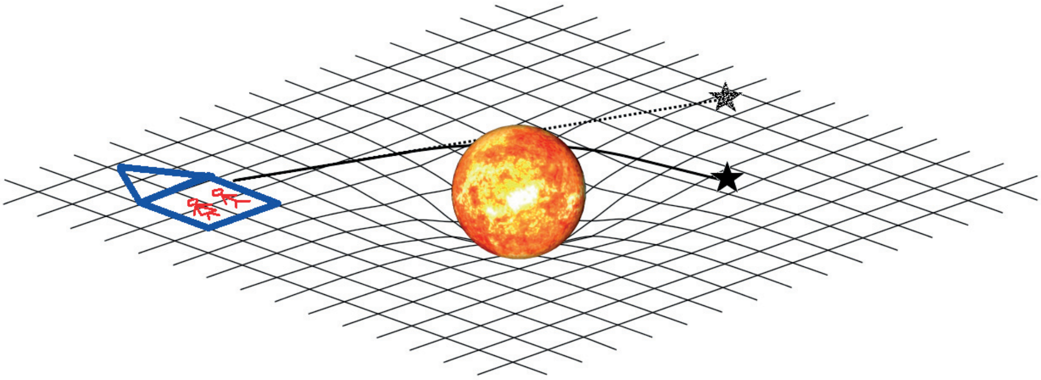


Figure 1.1: Illustration of the principle of a gravitational potential deflecting the trajectory of photons. The 2-dimensional space in which the two folks in the tiny house are living in is deformed by the presence of a massive star at the center of it. The solid line represents the trajectory of the light coming from the smaller black star on the right, that is bent when passing through the well, like would the trajectory of a marble in a bowl. However, the two folks in the tiny house, with their limited perception of space are unaware of this phenomenon. They thus believe that the star they observe is located along the dotted black straight line.

in the line of sight, a small difference was measured due to the bending of the light rays. At the time this observation took place, thanks to Einstein's explanations of the photoelectric effect (Einstein, 1905) it was accepted that the light had a dual wave-particle nature and behaved accordingly. Thus, based on these properties, GR predicted a well defined bending amplitude of the light rays. Interestingly, more than a century earlier Michell (1784); Soldner (1801) predicted that such a deviation should be seen as well, assuming the photons were particles subject to Newtonian mechanics. The latter predicted an angle of deviation twice smaller than the GR prediction. The bending amplitude observed by the 1919 observations reported in Dyson et al. (1920) disfavored the Newtonian framework prediction, yet corresponded very well with the prediction of GR, marking the first big success of Einstein's theory.

In the decades following this cornerstone observation, many predictions were made about what could in theory occur in space due to the bending of light rays. The one that interests us is when the perturbing body (also referred to as the deflector in the following, since it *deflects* the light) is massive enough, and the light source is located in a close to perfect alignment with the deflector and the observer. In such a case, the light rays going both sides of the deflector could converge towards a single point at the observer's position, from where the original source would be seen multiple times (Einstein, 1936). In such a case, the deflector would focus the light rays like an optical lens. This is why this phenomenon is nowadays called *gravitational lensing*, and the deflector is called the *lens*. Following the gravitational lensing equations, such a deflection that would produce multiple images of the lensed object is more likely to be observed when i) the deflector is very massive and ii) the source and deflector are both located far away from the observer. The former point can be easily addressed by considering as lenses objects much more massive than single stars, like nebulae or galaxies, as proposed by Zwicky

(1937a,b) for example. The latter point, however, requires the sources to be extremely bright in order to be seen by the observer. The distances at which gravitational lensing is at play being a few orders of magnitude higher than the size of galaxies, possible source candidates bright enough could then be, for instance, exploding stars (that are called *Supernovae*) or even other galaxies.

1.1.4 Quasars and lensed quasars

In a completely different field of Astrophysics, starting at the beginning of the 20th century, astronomers were observing in the sky peculiar compact objects emitting in the radio band. Their spectra were not properly corresponding to anything known so far but yet being somewhat similar to observed stars, they were incidentally named *quasars*, for quasi-stellar objects. Long story short (see Shields, 1999; Kellermann, 2013, and references therein for a review), it was admitted after much controversy in the early 1960's that these quasars were not stars but extremely bright, extra-galactic objects (Hazard et al., 1963; Schmidt, 1963). Nowadays, our current understanding of quasars indicates that most likely what we observe are regions full of very dense matter orbiting around a super-massive black hole located at the center of other galaxies, that emit tremendous amounts of light. For that reason, they are also commonly referred to as *Active Galactic Nuclei* or AGN.

In the scope of gravitational lensing, the most important point regarding quasars is that they are compact and very bright extra-galactic objects. Therefore, they represented excellent candidates to be lensed by galactic-size gravitational lenses. It took nonetheless more than another decade after the initial quasar observation to finally observe for the first time a gravitational lensing event. In 1979, Q0957+561, a close pair of two blue stellar objects that were looking extremely similar was serendipitously discovered. Quoting the discovery paper *"A less conventional view would find the quasars to be two images of the same object produced by a gravitational lens"* (Walsh et al., 1979; Walsh, 1989). Less conventional certainly, but nonetheless true.

1.1.5 Various regimes of lensing

Nowadays, we make the distinction between *strong, weak* and *micro* lensing regimes. These are all produced by the same principle that the light rays are bent by massive objects in space, and the distinction refers more to our ability to observe this effect.

We speak of strong lensing when multiple images are distinctly seen or when large arc-like features are displayed, as for the case of Q0957+561. Following this pioneer observation, other double, triple and quadruple-lensed quasars were discovered, as well as gravitational arcs. Arcs, first observed in the mid-1980's (Lynds and Petrosian, 1986; Soucail et al., 1987) result from a more precise alignment of the source and lens galaxy, where the lensed images are extensively stretched up, to the point they join each other. In case of a perfect alignment, the source is seen as a perfect ring around the lens galaxy, called the Einstein Ring. The first arcs observed were, however, not produced by the deflection from a single source but from a cluster of galaxies. In such a case, the complexity of the mass distribution in the lens cluster is sufficient to produce arcs for a larger range of source-lens-observer configurations. Figure 1.2 shows side by side the lensed quasar Q0957+561 and the arc in the galaxy cluster Abell-370 as seen when they were first observed, as well as more recent images of the same objects for comparison.

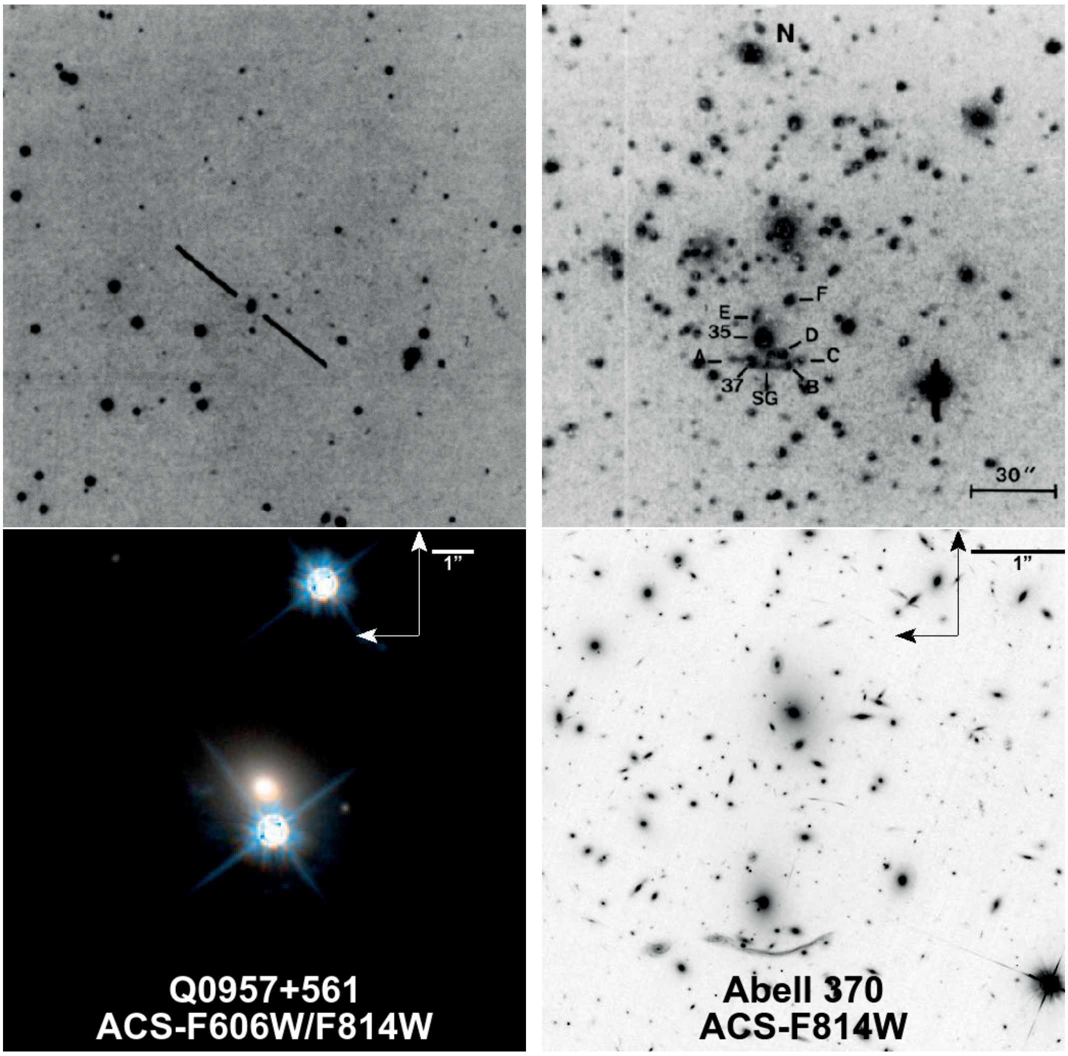


Figure 1.2: The top row presents Q0957+561 (left, Walsh et al. (1979) and Abell-370 (right, Soucail et al. (1987)), the first observed lensed quasar and gravitational arc seen at the time of their discovery. The bottom row presents current images of the same systems taken using the Hubble Space Telescope (credits: Kyle Stewart at www.masterlens.org)

Weak lensing is invoked when the lensing effect is too small to produce multiple images, and instead only a change in the shape of the observed object can be noted. It happens generally when the lens and source are not really aligned with the observer, yet still sufficiently close in projection and the lens sufficiently massive for the effect to be detectable. Weak lensing affect the ellipticity and alignment of the sources with respect to the lens. This effect is much more complicated to observe, since we cannot *a priori* distinguish if the potentially strange shape of a single object is due to its geometrical configuration relative to us, or due to weak lensing. It is only by looking statistically at the spatial distribution of ellipticities and alignment of many objects in a large field that the weak lensing effect can be detected. Thus, the first detection was made more than a decade after the first lensed quasar had been observed, when Tyson et al. (1990) showed that the galaxies in a large cluster tend to align perpendicularly to the direction of the cluster center. It took another ten years to observe the same effect on a wider scale, when (Bacon et al., 2000; Kaiser et al., 2000; Van Waerbeke et al., 2000; Wittman et al., 2000) independently measured a similar alignment caused by large-scale structures across cosmological distances, an effect called *cosmic shear*.

Finally, we speak of micro lensing when the deflector is so small that the only visible effect is to affect the brightness of the source. The idea that such effect could occur comes from Chang and Refsdal (1979); Gott (1981) and the term *microlensing* was first introduced by Paczynski (1986). In a sense, microlensing is similar to strong lensing, except that the lens is many orders of magnitude less massive. Regarding how difficult it was to observe a strongly lensed event in the 1980's, there were only few probable observable scenarios during which microlensing events could be witnessed. One plausible option (further discussed in Sec. 3.1.1) was that already strongly lensed images of quasars could be affected by stars in the lens galaxy crossing one of the images' position, as predicted by Chang and Refsdal (1979). On much smaller distances, microlensing could be used to detect binary companions to stars in the Milky Way (Mao and Paczynski, 1991), or even exoplanets (Gould and Loeb, 1992). Following estimates of possible galactic microlensing event that could be detected (see e.g. Paczynski, 1991), systematic observations of the Galactic bulge were launched by the Optical Gravitational Lensing Experiment (Udalski et al., 1992, hereafter OGLE). A third possible scenario invoking microlensing is that dark matter in the Milky Way halo, if existing in form of compact and massive particles should also affect the brightness of background sources, as reported by the MACHO (Alcock et al., 1993) and EROS (Aubourg et al., 1993) experiments.

It took the observers several decades after the theoretical predictions to finally observe gravitationally lensed systems in the sky, but in the meantime theoretical astrophysicists and cosmologists were also working on the practical applications of such discoveries, providing they could be made at some point in the future. In the scope of this thesis, certainly the most important prediction has been made by Sjur Refsdal, who suggested in 1964 that strongly lensed point-sources, thus seen multiple times, can be used in the framework of General Relativity to measure the rate at which the Universe expands (Refsdal, 1964b,a). Such a rate is what we call nowadays the *Hubble constant*, or H_0.

1.1.6 H_0 and the expansion of the Universe

At the time Refsdal suggested that Strong Lensing could help to measure H_0, it was commonly accepted that the Universe we are living in was not static, but was in expansion.

The theory of General Relativity allows a description of the whole Universe as being either static as in Einstein's original beliefs (Einstein, 1917), or expanding/contracting at a computable rate as independently pointed out by Alexander Friedmann (Friedmann, 1922) and Georges Lemaître (Lemaître, 1927). Lemaître, building on observations of recessional velocities of extra-galactic nebulae made by Vesto Slipher and Gustaf Stromberg (Stromberg, 1925) made the connection between the expansion of the Universe predicted in his theoretical models and the recessional velocity of the nebulae v that exhibited a dependency on their distance d relative to the Earth. Doing so, Lemaître was the first to propose a *value for the expansion rate of the Universe* based on observations. However, for reasons that now belong to history, it is only two years after Lemaître's pioneer work that the idea really kicked in through the work of Edwin Hubble, that basically redid Lemaître's analysis (Hubble, 1929). Hubble's paper provides, however, a nice illustration of the distance-velocity dependency in the form of a velocity diagram, reproduced in the left panel of Fig. 1.3. It shows a straight relation between these two observables in the form of $v = K \cdot d$, where the slope of the relation K represents the expansion rate. Hubble's paper having been somehow much more publicized and recognized than Lemaître's one at the time, the K coefficient ended up being renamed H_0, after Hubble. The 0 subset in H_0 denotes here that this measurement concerns the expansion rate as observed at the present time, or t_0. With our current knowledge, we now suppose that the expansion rate was greater earlier in the history of the Universe, but at the time Hubble performed his observations the nebulae he was interested in were not far enough from the Earth to test this particular fact. Thus, the "constant" term in the Hubble constant must be considered as a constant in space -we assume that we observe from a non-privileged spot in the Universe- but not in time.

This distance-velocity relation, when we think about it, is perturbing. It is easy to understand that the components of the Universe, stars, clusters, galaxies, etc... are not static. Massive objects are attracted to each other through gravitation and are always in motion with respect to each other. That motion is driven by an *attractive* force and at large scale, the celestial bodies should get closer to each other. What is observed here is that beyond a minimal distance at which the matter tend to effectively cluster, everything is moving *away* from everything else! Comparing the Universe with a chocolate muffin being baked in the oven, the current expansion of the Universe can be compared to the muffin being heated and thus inflating. Two chocolate chips in the middle of it would see themselves moving away from each other, driven by the inflation of the whole muffin. In this simple analogy, what drives the expansion would be the heat provided by the oven. But when considering the Universe, the nature of what drives the accelerated expansion, called *dark energy* nowadays, is still a mystery.

Quantifying the expansion rate requires to be able to measure accurately the absolute distance

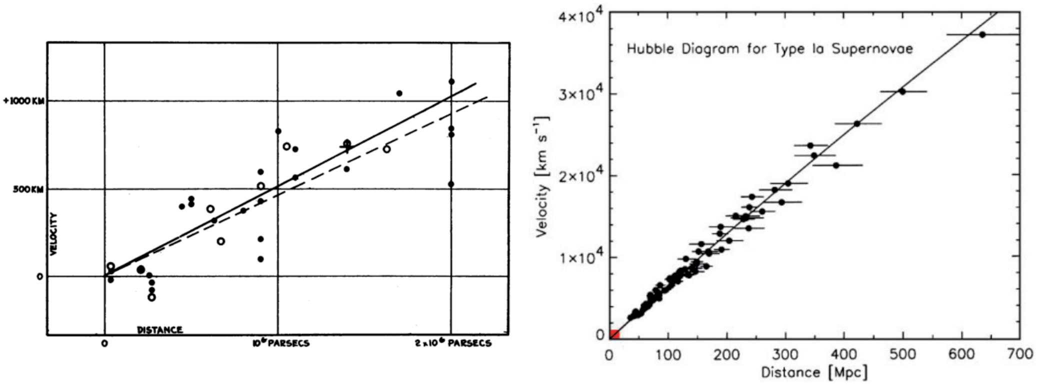


Figure 1.3: The left panel presents the relation between the observed escape velocity and distance of extra-galactic nebulae, as originally presented in Hubble (1929). Black points are individual measurements and open points are averages over multiple measurements binned by distance, fitted by the solid and dashed lines, respectively. The slope of each line gives a value of the Hubble constant H_0. The right panel presents the same relation using type-Ia Supernovae acquired until 2002 (reproduction from Kirshner (2004)). The small red square at the bottom left corresponds to the diagram from 1929 on the left panel.

d mentioned above. Measuring *absolute* distances in astronomy is far from being easy, as we will see later on, and also not always necessary. It is much more simpler to use a *relative* scale to order the position of objects in the Universe. Taking advantage of the fact that objects located further away from us are moving away faster due to the expansion of the Universe, we can define a scale based on how their color as perceived from our position in space is altered by their recessional motion - some kind of equivalent to the Doppler effect for sound waves, that we call *redshift*. The farther an object is from us, the redder it appears. The redshift can also be used as a relative time scale; the further away we look, the more into the past we look at since the light travels at a finite speed.

Coming back to the expansion of the Universe, the mathematical term in the equations of General Relativity describing the presence of that mysterious driving force is denoted by the Greek letter Λ, and is referred to as the *Cosmological constant*. What Lemaître theorized and confirmed with his observations was that the Universe is not static, as believed by Einstein, nor shrinking as it would likely be in the absence of large-scale forces other than the gravitation, but is instead expanding. What was not clear at that time was how this expansion will evolve over time. Attractive matter competes against a repulsive Cosmological constant; if the Universe is currently expanding, it meant that the Cosmological constant is winning over Gravitation. Yet, the absolute expansion rate should diminish over time, since the Universe gets bigger and bigger. Multiple scenarios were suggested as in Eddington (1930); Lemaître (1931a) for instance, where the Universe would either ending up collapsing on itself, reach a steady state or expand indefinitely.

The case of an infinite expansion was, somehow, the least appealing of the three possible outcomes. In the decades following the results of Lemaître and Hubble, the consensus was that the expansion should be decelerating, and many experiments and observations were imagined and devised in that regard (Bondi and Gold, 1948; Hoyle and Sandage, 1956; Sandage, 1961; Schmidt et al., 1998). The present-time deceleration parameter q_0 was introduced, emerging from the GR equations and one important success for cosmology at the time would have been to precisely determine its value, and ensure it is indeed positive - by definition, $q_0 > 0$ implies a decelerating universe. Although a few early hints were provided towards an opposite conclusion (Efstathiou et al., 1990), the vast majority of the scientific community was taken aback when two independent groups, using observations of type-Ia Supernovae pointed towards an *accelerating* expansion of the Universe (Riess et al., 1998; Perlmutter et al., 1999)! Indeed, since looking away from us corresponds at looking in the past because of the finite speed of light, measuring the distances and recessional velocities of far away objects - in the case of Riess et al. (1998) and Perlmutter et al. (1999), exploding stars called Supernovae located in other galaxies - allowed us to directly probe the evolution of the Universe expansion at different epochs. Similar to the Hubble diagram presented earlier, Kirshner (2004) presents the relation between distance and velocity of type-Ia Supernovae, reproduced on the right panel of Fig. 1.3. The deviation from a straight line is barely visible, but the evidence is there: the Universe was expanding *slower* in the past, which means the expansion is currently *accelerating*!

1.1.7 Observing the Universe as a whole

This game changing measurement, for which Adam Riess, Saul Perlmutter and Brian Schmidt were awarded the Nobel Prize of Physics in 2011, made the Universe even more puzzling to cosmologists. It was not in contradiction with the theory of General Relativity, but conceptualizing something that is *accelerating* the expansion proved to be quite a troubling task, to say the least. One thing known for sure was that this "energy" was not directly visible by any means - which earned it its current designation of *dark energy*. Nowadays, scientists of many different fields of physics are trying to solve the mystery of dark energy, but despite their best efforts only little progress has been achieved. In that regard though, the Hubble constant might play a crucial role. H_0 is intrinsically linked with the dark energy, as the expansion rate is directly affected by the amount of repulsive energy in the cosmos.

In this context, it is important to understand that an independent measurement of the Hubble constant alone is not that useful; it is only a piece of the big puzzle that is the modern day cosmology. If we want at some point to have the chance to completely understand the Universe, one needs not only to look at it through independent, specific lenses (no pun intended), but to consider it as a whole. As said before, if the dark energy is responsible for the expansion, it still has to compete against ordinary matter that is attractive, and thus would slow the expansion. And matter, to the great surprise of all, is not only composed of the ordinary particles that makes human beings, planets and stars but also of another type. The latter has so far not been seen interacting with the ordinary matter nor with itself in other ways than through gravitational interactions, hence its current denomination as *dark matter*. Hints of its existence have been around since the first half of the 20th century to explain the observations of stellar motions in the Milky Way (Kapteyn, 1922), until being indirectly observed by looking at the motion of whole galaxies evolving in larger clusters (Zwicky, 1933), as well as in single isolated galaxies a few decades later (Rubin and Ford, 1970). But beyond the existence of dark matter and its impacts on cosmology, the important point here is to understand that whatever mysterious, dark or ordinary component we consider, they all are at interplay. They act together to give the Universe its current shape and properties.

Another remarkable theoretical prediction of the first half of the 20th century, once again rooted in the pioneer work of Lemaître (1931b), is the existence of a starting point for our Universe where both space and time were born, nowadays called Big Bang. Since the Universe is in expansion, it might be that at some point in the past it was infinitely smaller as well as extremely dense and hot in order to contain all the energy and matter nowadays distributed all across the Universe. Right after its birth, the Universe has undergone a short and extremely fast expansion phase called the inflation, followed by a second, slower but steadier expansion phase during which it gradually cooled down. The photons emitted when the Universe was hot were rapidly absorbed through interactions with other particles, but once the Universe reached a state large and cool enough, the photons had suddenly enough space to roam freely without interacting. Such photons, if able to travel through space and time up to us today, would provide us with the earliest known picture of the Universe as well as a confirmation that the Big Bang model is correct.

By trying to understand the observed abundances of elements in the Universe, Gamow (1948); Alpher and Herman (1948) were predicting the wavelength at which this first generation of photons would be seen today. In 1964, Arno Penzias and Robert Wilson serendipitously observed these photons (Penzias and Wilson, 1965) at a wavelength close to the predictions made two decades earlier. This measurement once again deeply changed the way the scientific community regarded cosmology. It was the first measurement of the Cosmic Microwave Background (hereafter CMB), considered as a strong observational evidence in favor of the Big Bang model and for which Penzias and Wilson were awarded the Nobel Prize in Physics in 1978.

One of the great success of this measurement is that it launched a new era of *precision* cosmology. In the wake of the CMB discovery, it was suggested that with a better resolution, anisotropies could be observed in it (see e.g. Harrison, 1970; Peebles and Yu, 1970). Indeed, since the CMB photons were emitted from everywhere in the Universe at some point in the past, they are currently visible from every direction in the sky. And depending on what were the Universe's composition and properties in the regions that emitted these photons billions years ago, one could see tiny shifts in their wavelengths, depending on where exactly they come from. Such anisotropies would be a relic signal of the Universe's composition in its early life - a signal that is thus an extremely valuable information to understand the Universe's history and composition.

Satellites as well as ground-based and high atmosphere balloon-based instruments were devised to observe these anisotropies. Among them, certainly the most publicized ones were the satellites *Cosmic Background Explorer* (COBE Smoot et al., 1992a), *Wilkinson Microwave Anisotropy Probe* (WMAP Bennett et al., 2003) and *Planck* (Planck Collaboration, 2005) which delivered in the past 25 years more and more accurate maps of the CMB, as illustrated in Fig. 1.4. The color scales are not reproduced in this figure, but in all three cases the variations across the map were extremely tiny, corresponding to a variation of temperature in the emitting regions of no more than half a degree.

The study of the spatial distribution of the anisotropies in the CMB maps is an extremely powerful tool to constrain the Universe. To properly model the statistical properties of such maps, one needs to take into account all the parameters of the considered cosmological model, that are all at interplay. A statistical study of the CMB allows to constrain jointly the value of all these cosmological parameters. The most recent CMB measurements from Planck have given rise to the most stringent constraints on the cosmological parameters up-to-date (Planck Collaboration et al., 2016c).

Then, the question is why should we put so many efforts into measuring only the Hubble constant, when CMB observations are able to deliver such stringent joint constraints on the whole cosmological parameters ? The answer resides in the question itself. The constraints on the cosmological parameters from CMB are not independent from each other, and they all rely on initial assumptions that still have to be firmly established. And among them, without much surprises, is the nature of dark energy: shall it behave slightly differently from what we expect, then the whole shape of the puzzle would change. In that regard, independent measurements of H_0 or any other cosmological parameters are to be seen as a complementary way to assess if our current picture of the Universe is correct. A naive, yet evocative way of picturing this complementarity is to imagine the Universe as a simple three-dimensional object, like the cork presented on Fig. 1.5. As observers, the only information we have access to is the shadow of the cork, representing our results. The shadow is casted by a source of light shining from a given perspective representing one experiment. A single experiment corresponding to a single perspective could be as precise and accurate as possible, it will only give us access to part of the information. Complementary experiments will cast multiple shadows and only then we will be able to fully grasp the nature of our Universe.

1.2 The use of lensed quasars in cosmology

To measure the expansion speed of the Universe and thus constrain the cosmological models, the method suggested by Refsdal (1964b,a) using single galaxies as deflectors looked promising thanks to its apparent simplicity. The strong lensing phenomenon presents two interesting properties. First, the lensed images could be *magnified*, i.e. seen brighter than they would appear without being lensed. Second, the luminosity variations of the lensed images reflects the variations of the source, but are observed with a *time delay*, due to the difference in path

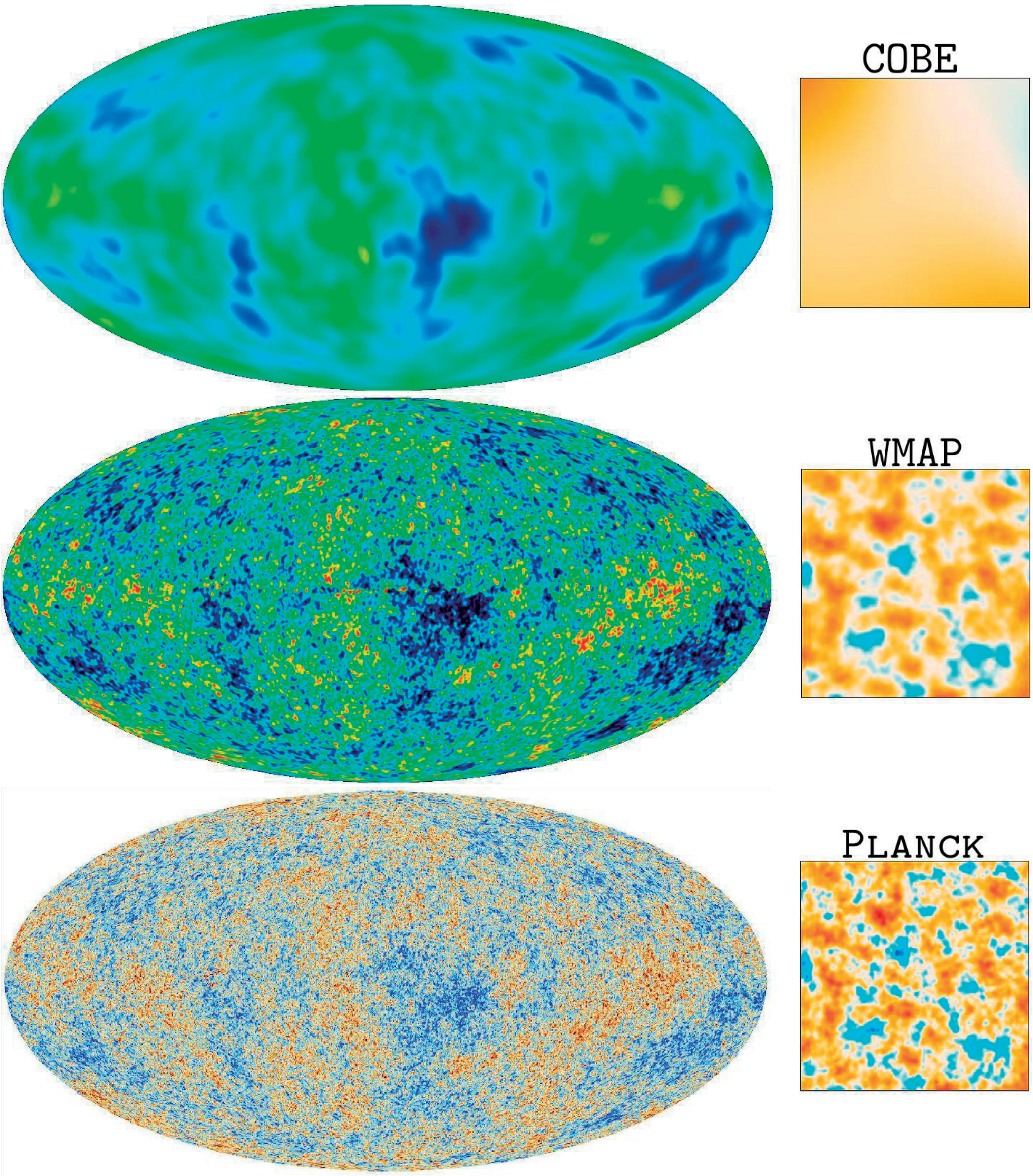


Figure 1.4: Evolution of the resolution of the Cosmic Microwave Background maps, as measured by COBE (Bennett et al., 1996, top), WMAP (Bennett et al., 2013a, middle) and Planck (Planck Collaboration et al., 2016a, bottom). These maps are a 2d projection of the CMB radiation measured on every direction in the sky, where a change of color indicates a variation in the measured temperature. The color scales (different for each map) are not reproduced here . The squared inserts on the right represent a similar region seen on each map, to highlights the improvement in spatial resolution. Credits: NASA / JPL-Caltech / ESA.

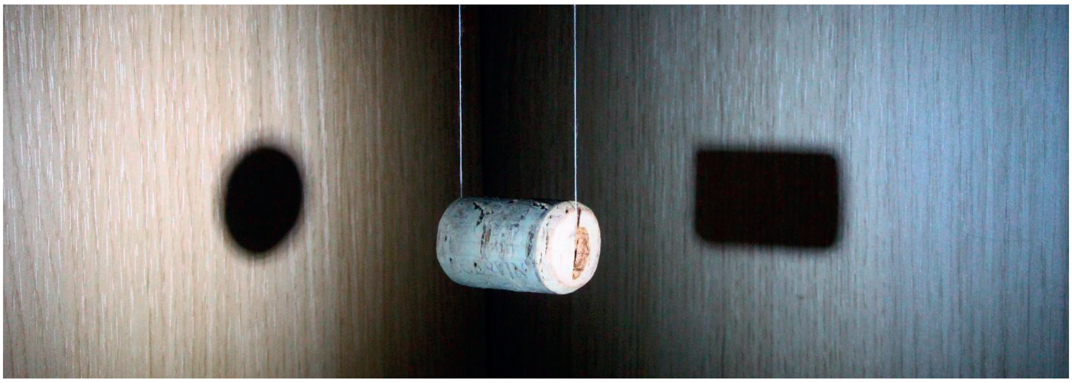


Figure 1.5: Allegory of the multiple experiments currently performed in cosmology to understand the nature of the Universe. A cork, representing the Universe, is illuminated by two sources of light from two different angles, symbolizing two complementary experiments. The goal of each experiment is to understand the shape of the object they are shining at. The result of the individual experiments are represented here by the two shadows casted by the light. Looking at the results individually, one might think that the cork's shape is either a rectangle or a sphere. It is only when we consider the two results together that the true shape of the cork becomes known. Picture adapted from Szabo (2014).

lengths of the photons. These properties, correctly used in the framework of GR, could lead to a direct measurement of H_0. *Time-delay cosmography* was born .

As mentioned earlier, the first main difficulty of the technique was to find sources bright enough to be observed from the earth. Following the discovery of the first lensed quasar Q0957+561 by Walsh et al. (1979), the interest for this technique received a tremendous boost. As of today, more than 140 scientific publications were at least partially dedicated to the study of Q0957+561, and many more were dedicated to the study of gravitational lensing and time-delay cosmography in general. But despite the growing interest for lensed quasars, their small apparent size made their search complicated. As a result, the first lensed quasar discoveries following the initial often happened serendipitously. Among the early lenses discoveries, let us mention PG1115+080 (Weymann et al., 1980) as the first triple-imaged quasar, Q2237+0305 also called the Einstein Cross (Huchra et al., 1985) as the first quadruple-imaged quasar, HE1413+117 also called the Clover Leaf (Magain et al., 1988) as another quadruple-image quasar with a large separation (larger than 1") between the four images and B1608+656 (independently discovered by Snellen et al., 1995; Myers et al., 1995, in radio-band and visible-band, respectively). In the 1990's, multiple sky surveys designed to find gravitational lenses were initiated. The *Jodrell Bank/VLA Astrometric Survey* (JVAS, Patnaik et al., 1992), the *Cosmic Lens All-Sky Survey* (CLASS, Myers et al., 1995), or the Hamburg/ESO Survey (HES, Reimers et al., 1996; Wisotzki et al., 1996a,b) were all designed to significantly improve the number of lenses of various configurations.

The idea from Refsdal (1964a) originally proposed to use lensed Supernovae to constrain H_0, although it appeared much more convenient to use lensed quasars after they have been discovered. Regardless of the nature of the source, the method requires that two observable quantities are measured as precisely and accurately as possible. First, the mass profile of the lens galaxy (i.e. the total mass enclosed at a given radius as well as the mass profile at various radii) and second, the time delays between the lensed images. We speak of a *time delay* between two lensed images if their variation in luminosity - which are the intrinsic variations of the quasar source - are not observed simultaneously in the two images. The delay between these observations is due to the different length of the paths traveled by the photons as well as to the gravitational potential of the lens galaxy which slows the photons down - an effect know as the Shapiro delay. In theory, a measure of the time delay "simply" requires to monitor the luminosity variations of the quasar images over time. In practice though, the precision required by the observations is far from easy to achieve.

1.2.1 Measuring time delays: the controversy of Q0957+561

Right after the discovery of the first lensed quasar Q0957+561 in 1979, monitoring campaigns were set up in order to measure the time delay between the two lensed images. In that regard, Q0957+561 is a good candidate. Both images are really bright (magnitude ~17 in R band) and easy to resolve (separation > 6 arcsec). The lens galaxy is, in projection, much closer to the brightest image, making the whole system asymmetric and thus ensuring a long time-delay (Schneider, 1985). During the 1980's, many predictions and wild guesses from sparse observations were announced. Yet, the first solid result from a monitoring campaign was proposed by Vanderriest et al. (1989). Following a monitoring campaign of seven years in the optical band reproduced in the left panel of Fig. 1.6, the authors found a delay of approx. 415 days. One year later, this result was confirmed by Schild (1990), following a ten-years monitoring campaign with various telescopes. But two years later, yet another monitoring campaign in the radio band yielded a delay of ~520 days (Lehar et al., 1992), a value corrected to 540 days by Press et al. (1992b) through a reanalysis of the previously published data sets. After refinement of the optical data sets (Schild and Thomson, 1995), a new analysis by Pelt et al. (1996) yielded again a smaller delay of 420 days, close to the initial estimate. It is finally in 1997, 18 years after the discovery of Q0957+561 that Kundić et al. (1997) were able to solve the controversy. By monitoring the lensed images with an higher cadence, they observed a strong magnitude variation event in the two images that finally discarded the ~520 days delay and confirmed the original prediction of Vanderriest et al. (1989). The event is reproduced in the right panel of Fig. 1.6.

The goal then - measuring a time delay between the luminosity variations of two lensed images - appears straightforward to solve. After all, it simply consists of recognizing a similar pattern in the two light curves and identify the time shift between two occurrences. But as often, the real difficulty lies in the practical applications. As stated in Kundić et al. (1997) when talking about the previous results from Press et al. (1992b); Pelt et al. (1996), *"These two rough values* (i.e. the two estimates of the time delay in Q0957+561) *have been obtained both by applying the same statistical techniques to different data sets and by applying different statistical techniques*

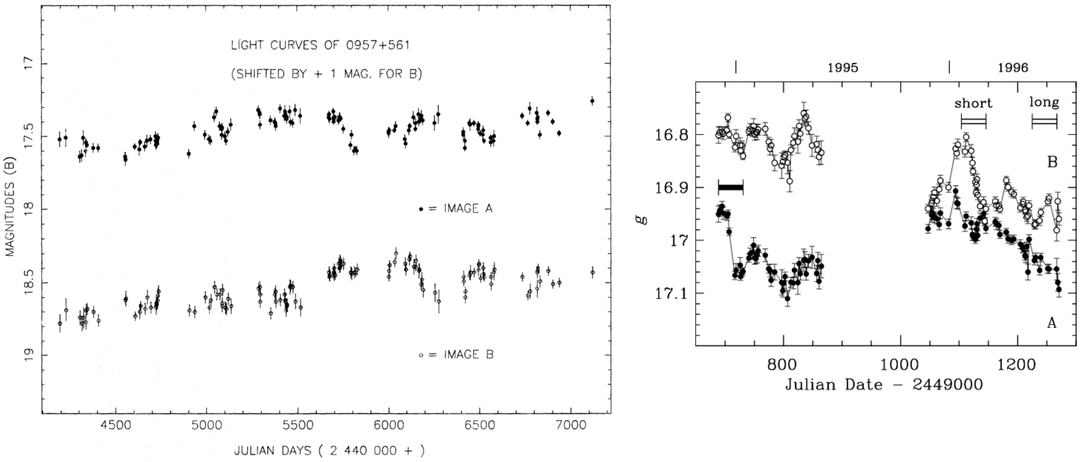


Figure 1.6: The left panel presents the light curves of the double lensed quasar Q0957+561 in the R band after 7 years of monitoring campaign, reproduced from Vanderriest et al. (1989). The right panel presents the light curves of the same object, reproduced from Kundić et al. (1997). The black horizontal bar indicates the strong variation in luminosity of the image A. The two white horizontal bars indicates where this event should be observed again in the image B, if the time delay is of ~420 days (short) or ~520 days (long). The data strongly speak in favor of the short delay.

to the same data". That quote highlights the main difficulties faced when measuring time delays, that can be summarized in three points:

1. A high-quality photometric precision is usually mandatory. Since the goal is to observe variations in luminosity of the images, the photometric error should *at least* be an order of magnitude smaller than the luminosity variations for the observed structures in the light curves to be well defined. A purely empirical yet very practical rule-of-thumb is that if the time delay is not visible by eye in the light curves, then it is unlikely that numerical recipes will yield a solid estimate. The observations of Vanderriest et al. (1989) (left panel of Fig. 1.6) illustrate well this issue. In many seasons, no structure are visible at all due to the poor photometric precision[1].

2. A regular sampling -if possible at high cadence- is extremely important as well. The luminosity variations of the quasar tend to statistically follow certain trends (Mushotzky et al., 2011), yet are somehow unpredictable at small time scales: the quasar can undergo a phase of quiescent evolution for a few month and suddenly exhibit a burst of luminosity that makes an extremely valuable feature when it comes to the determination of the time delay. Since these features can be quite short, observations with regular sampling are mandatory to catch them. The observations of Kundić et al. (1997) (right panel of Fig. 1.4) are an excellent illustration of this fact. Without the weekly cadence monitoring

---

[1] Let us note that despite this fact, Vanderriest et al. (1989) measured a correct delay. Yet it took eight extra years to the community to validate these results.

used then, it would not have been possible to lift the controversy on the time delay of Q0957+561.

3. Finally, a long baseline of observations is required in order to minimize the effect of the extrinsic variability. Such effects might occur, for example, because of stars in the lens galaxy that affect independently the observed luminosity of the quasar images - this is called microlensing, detailed in Sec. 1.1.5 and 3.1.1. These extrinsic variations that occur unpredictably affect independently the lensed images. When combined with the effect of irregular sampling and photometric noise, it becomes difficult to distinguish between intrinsic and extrinsic luminosity variations. The year Q0957+561 was discovered, Chang and Refsdal (1979) estimated that it might be affected by microlensing. A decade later, Schild and Smith (1991); Falco et al. (1991) showed that it was indeed certainly the case - an assumption that was confirmed shortly after the uncertainty on the time-delay has been lifted (Pelt et al., 1998; Refsdal et al., 2000; Wambsganss et al., 2000).

Even seasoned observers are often tricked by at least one of the difficulty mentioned above, eventually making bold - if not wrong - claims about time-delay measurements and plaguing the field with incorrect results. The pressure is yet understandable: the observations require a huge amount of telescope time, making failure of measuring something meaningful hardly an option. That is why a proper formalism and measuring techniques must be used. Such have been established by different research groups and collaborations, including COSMOGRAIL - the one I am part of. These developments will be presented in more details in Chapter 3.

1.2.2 The lens galaxy and its surroundings: between assumptions and observations

Even if we assume that the time delays are sufficiently well measured, we need to add another ingredient to the mix if we want to measure H_0: the lens galaxy. To continue with the analogy of the optical lens introduced earlier, we understand that the bending of light rays by an optical lens depends on its physical properties, namely what is it made of, and how thick and curved is the lens at the position where the light rays cross it. Similarly, in the case of a gravitational lens we need to know what is the lens mass distribution and where are the lensed images located with respect to the center of the lens galaxy. If the latter can be precisely measured with high-resolution images of the lens galaxy, the former is more complicated to obtain. It requires either very specific high-quality observations, strong assumptions on the nature of the lens or a bit of both.

The mass distribution of the lens galaxy is assumed to be reasonably well described by an analytical profile. But a profile tells us only about the shape of the lens galaxy, without informing us on an absolute quantity related to it. In other terms, what needs to be determined are the shape of the mass profile, and the total mass of the galaxy - or the total mass enclosed within a given radius. The former can be either inferred from numerical simulations or observations of the luminosity profile. The latter requires either the measurement of the velocity dispersion of the stars in the galaxy - statistical properties of their velocity distribution being related to the total mass of the system - or, in the few lucky cases where an Einstein ring is seen around the lens galaxy, the size of the ring that is related to the total mass "inside" it.

On top of that, another thing should be considered: the source-lens-observer system lies in the middle of a huge web of galaxies that might have a considerable impact on the measurements. First, the lens galaxy itself might be part of a larger cluster, whose other galaxies might impact the bending of the light rays. Falco et al. (1985); Gorenstein et al. (1988) and more recently Schneider and Sluse (2013) showed that such effects, called mass-sheet transformations are extremely tricky, leaving all the observables but the time delays and the absolute magnification of the images unchanged. The measured time delays could in principle be used to discriminate between two very different mass models that predict in turn discrepant time delays, but that would work only when the predictions are very different. The absolute magnification of the images can be known only if the absolute luminosity of the source is known, which is not the case for quasars. Second, the other galaxies along the line-of-sight have a small but yet potentially significant impact on the light rays, focusing and defocusing them all along their path and introducing a bias in the estimates made from the observables (Keeton and Zabludoff, 2004).

1.2.3 H_0 estimates from Q0957+561

In theory, with solid time-delay estimates as well as strong constraints on the lens and its surroundings, one should be able to determine the Hubble constant. Let us have a look at a practical example by going back once again to Q0957+561. Young et al. (1980) showed that Q0957+561 was located in a rich cluster of galaxies, whose effect was treated in Falco et al. (1985) who put an upper limit of ~90 $\mathrm{km\,s^{-1}\,Mpc^{-1}}$ on H_0. Using a time delay of ~420 days (i.e. the value finally accepted after the observations of Kundić et al. (1997)), Rhee (1991) found H_0 =50±17 $\mathrm{km\,s^{-1}\,Mpc^{-1}}$. A few years later, Grogin and Narayan (1996) derived a more precise value of H_0 =$82^{+7.2}_{-3.0}$ $\mathrm{km\,s^{-1}\,Mpc^{-1}}$, whereas Kundić et al. (1997) found H_0 =64±13 $\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and Falco et al. (1997) had a hard time to decide between H_0 =62±7 $\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and H_0 =67±8 $\mathrm{km\,s^{-1}\,Mpc^{-1}}$ due to the uncertainties on the lens stellar velocity dispersion. Oscoz et al. (1997) opted for H_0 =64^{+14}_{-15} $\mathrm{km\,s^{-1}\,Mpc^{-1}}$ whereas Bernstein and Fischer (1999) were more conservative, advocating for H_0 =77^{+29}_{-24} $\mathrm{km\,s^{-1}\,Mpc^{-1}}$ at a 2σ confidence. All these estimates, along with a few other, are reproduce in Fig. 1.7 below.

We can reasonably imagine that all these authors did their best to be as accurate as possible in their whole analysis. Then, why are the error bars sometimes so large and why some of the proposed values are in tension? Certainly because Q0957+561 -and more generally the whole time-delay cosmography analysis- is a perfect illustration of the science case that is easy to solve in theory, but that becomes actually much more complicated when it comes to practically measure things. The technique from Refsdal (1964a) applied to quasars is relatively simple and mathematically elegant. Yet, considering all the possible issues mentioned above

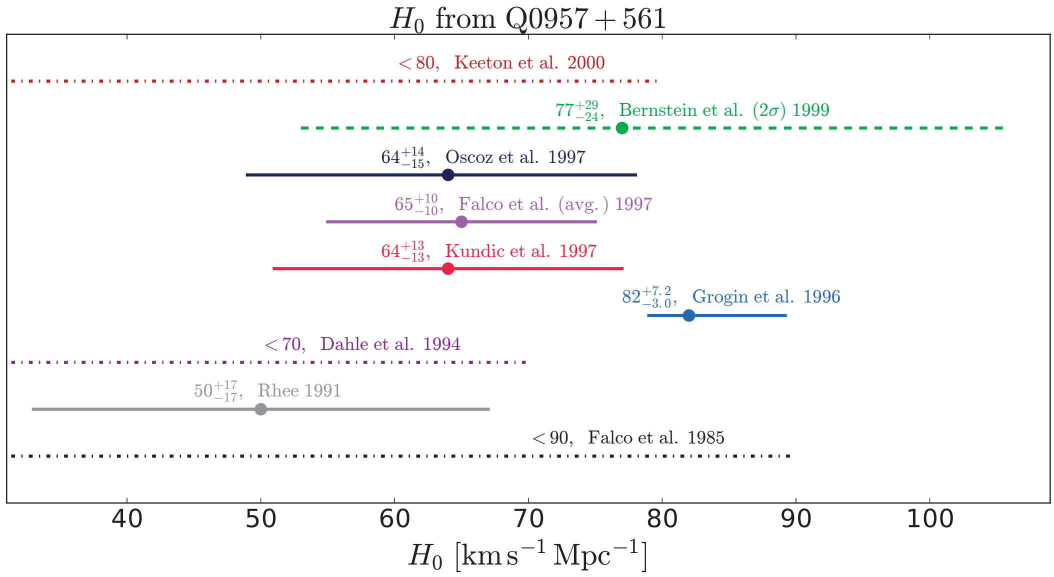


Figure 1.7: Non exhaustive H_0 estimates from Q0957+561 published between the years 1985 and 2000, only based on the admittedly correct value of the time delay measured by Kundić et al. (1997). The solid lines represents 1σ uncertainties, the dashed line from Bernstein and Fischer (1999) represents a 2σ uncertainty and the dot-dashed lines represent upper limits.

that could arise in the time-delay measurement and in the lens modeling, it leads to a not-so-precise determination of H_0, or worse, a wrong estimate. This happens because some sources of uncertainty have been ill-defined or even neglected. Even with the most careful efforts, effects like the microlensing on the mass-sheet transformation are complicated to properly account for. And sometimes, the degeneracies inherent to the observations cannot be lifted, but only carefully introduced in the total error budget of the measurement of H_0.

These difficulties should, however, not be seen as a no-go for time-delay cosmography. First, all the methods aiming at measuring H_0 are subject to uncertainties and sources of error to some extent - such will be presented in more details in Sec. 1.3. Second, the new instruments and telescopes developed over the years provide more and more precise observations, bringing new stringent constraints and thus reducing the assumptions we have to make, lifting *de facto* some sources of uncertainty. For example, a good quality imaging of the lens galaxy provides a luminosity profile, tracing the spatial arrangement of luminous matter and determining with great precision the center of the lens galaxy. Third, more and more lensed quasars are discovered years after years, some of them being perfectly suitable to time-delay cosmography. Very importantly, the lensed images configuration and the lens galaxy neighborhood can change drastically between two targets. Thus, some targets are more suitable than others when it comes to precisely measure the time delays, or to minimize the effects of the environment of the lens galaxy. Since each single lensed quasar analyzed can yield a value for H_0, a large sample of lensed quasars will yield multiple values that can be combined into a more precise one.

1.2.4 A cosmological Swiss Army knife

Like a proper cosmological Swiss Army knife, strong gravitational lensing has many interesting cosmological implications. Blandford and Narayan (1992); Wambsganss (1998); Meylan et al. (2006) give nice reviews of the applications of gravitational lensing in general, and more recently Treu (2010) did the same with strong gravitational lensing in particular. Although my goal is not to give an in-depth review of the astrophysical and cosmological applications of gravitational lensing, it will be a little bit unfair not to at least mention the other wonders that can be achieved by strong lensing.

- The strong lensing effect has the admirable property to stretch and magnify the source, if the geometrical configuration is favorable. Of course, studying quasars do not require for them to be lensed. Yet, a lensed source is seen in greater details, or becomes simply bright enough to be properly studied whereas without being lensed it would have been too faint to be observed. In such cases, the lens galaxy acts like a natural telescope that allows us to observe whatever lies behind. For example, Chartas et al. (2003, 2007) observe a relativistic outflow in PG1115+080, revealing information about the kinematics of the system and the formation of the quasar host galaxy; Marshall et al. (2007) is able to resolve the small scales (sub-kiloparsec) of a quasar host galaxy that appeared stretched; Refsdal et al. (2000); Eigenbrod et al. (2008); Poindexter et al. (2008); Chartas et al. (2009); Sluse et al. (2011) are able to constrain the accretion disk structure by comparing how the lensed images are differently affected by microlensing; using microlensing again on 11 different lenses, Poindexter and Kochanek (2010a) determine the usual orientation of the quasar accretion disk; more recently, Ding et al. (2017a) disentangle the light from the quasar and its host galaxy and studies the correlation between their luminosity.

- The extrinsic variations of luminosity of the lensed images provoked by perturbing bodies in the lens galaxy, i.e. microlensing, offer a direct view on the lens galaxy content at the position of the lensed images. Indeed, the frequency and amplitude of microlensing events are linked to the abundance and velocity of these "micro deflector". Stars and/or hypothetical compact dark matter objects create short luminosity variations in a light curve, whereas dark matter sub-halos that are supposedly more spatially extended affect the images on a longer timescale. By looking at the microlensing in Q2237+0305, Poindexter and Kochanek (2010b) estimate the mean mass of the stars in the lens galaxy. Mao and Schneider (1998); Dalal and Kochanek (2002) find that microlensing events attributed to large substructures in the lens galaxy halo predict a given amount of sub-halos, that has been predicted to be much more numerous in numerical simulations (Klypin et al., 1999; Moore et al., 1999) - a problem that so far has not been solved (Kravtsov, 2010)[2]. Similarly, the low abundance of short-term events can be used to put constraints on the nature of dark matter (Wambsganss et al., 2000;

---

[2] Amusingly, observers refer to this issue as the "excess of sub-halos problem" where theorist prefer speaking about the "missing satellites problem"

Schechter and Wambsganss, 2002), as recently demonstrated in Mediavilla et al. (2009, 2017) who strongly limited the possibility of black holes as dark matter particles.

1.3 A brief history of H_0 measurement techniques

The Hubble constant has been estimated for the first time not by Hubble himself as one may think, but by Lemaître (1927) two years prior Hubble's publication (Hubble, 1929). Using the recession velocities of nebulae as the one presented in Fig. 1.3, Lemaître derived a value of K=~ 625 km s^{-1} Mpc^{-1}- the constant was obviously not called H_0 in Lemaître (1927); Hubble (1929). Following that seminal estimate, others attempts followed and refined the measurement towards a lower value. Fig. 1.8 presents the evolution of the Hubble constant since its initial measurement in 1927.

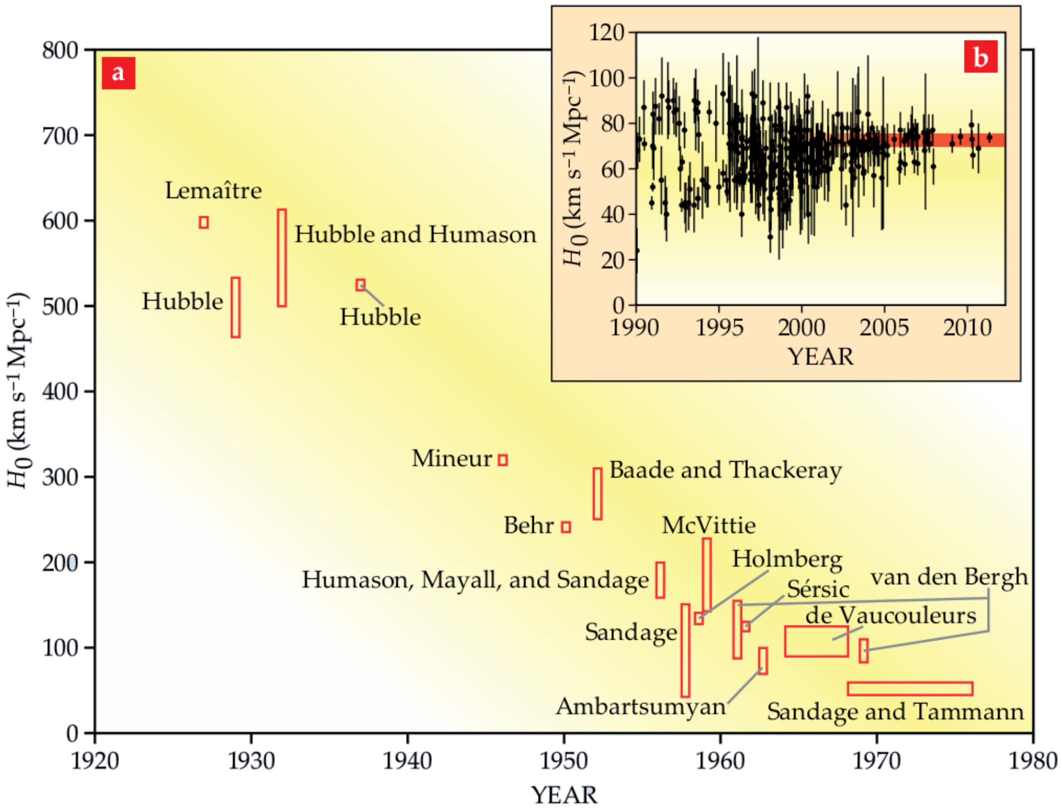


Figure 1.8: Evolution of the Hubble constant from 1927 to 2012, reproduced from Livio and Riess (2013). The large panel, labeled "a" presents the early estimates between 1927 to 1980. The insert, labeled "b" presents more recent estimates from 1990 to 2012. The red band indicates the 70-75 km s^{-1} Mpc^{-1} region where the most recent estimates tend to cluster.

Most notably, the large panel "a" shows that the estimate of H_0 strongly decreased over time. From several hundreds km s^{-1} Mpc^{-1} at the beginning of last century, the accepted value nowadays decreased to ~ 70 km s^{-1} Mpc^{-1} (small panel "b"). If the estimates were scarce at the beginning, they became more and more numerous over the years. It can be noted that the number of estimates per year peaked in the late 1990's and early 2000's. According to Bethapudi and Desai (2017), an highly disputable paper yet somehow apparently reasonably up-to-date when it comes to listing other publications before 2016, there has been close to 600 published estimates for the value of H_0.

Interestingly, the evolution over time shows us that most often, older estimates do not resist to the emergence of newer ones, in the sense that the more recent values do not lie within the error bars of the older ones. This might seems to be nothing too important, but there is certainly a lesson to learn from this. Not really often are the scientists conservative when it comes to publishing their findings and even so, despite their best efforts, their work might be subject to unknown sources of bias unaccounted for in their analysis. There can be many reasons for two measurements to disagree. One or both experiences can be biased, the underlying physics can be more complicated than one's think, or simply the tension might come from the way error bars are computed and interpreted, what is called a statistical fluke.

I already established earlier the need of independent measurements of cosmological parameters as a way to properly grasp the essence of the Universe. But simpler than that, calling for many independent experiments is a way to protect ourselves as much as possible from the unknown biases and sources of error mentioned above. Regarding the Hubble constant, there are currently three techniques that are precise enough to play that role in the current research context, and a few other ones promising enough that might blossom in a few years' time (see e.g Weinberg et al., 2013, for a recent review). I will conclude this introduction by briefly describing these methods, listing some pros and cons for each of them. I chose here to focus on the different technique without reporting their measurement of the Hubble constant. The comparison and discussion of the measured H_0 values is postponed to Sec. .

1.3.1 The distance-luminosity relations: a cosmic ladder

Inferring the distance of a source by measuring its luminosity is certainly the most famous and documented technique used to measure H_0 directly, and was also the technique originally used by Lemaître (1927); Hubble (1929). It consists of observing *standard candles*, i.e. objects with known absolute luminosity, and by comparing their relative luminosity compute their "luminosity" distance. In addition, measuring their redshift gives us a direct information on their recessional velocity. The slope of the relation linking the luminosity distance to the recessional velocity - assumed constant at small scales - is the Hubble constant H_0.

There are two main difficulties with this technique. First we need to be sure to effectively measure recessional velocities that are due to the Hubble flow, i.e. the expansion of the Universe. Peculiar motion of the standard candles in their host galaxy, or motion of the galaxies themselves in larger clusters can bias the measurements (Tammann and Leibundgut, 1990; Hui and Greene, 2006). Second, the assumption that the standard candles are indeed standard, and that the zero point of their distance-luminosity relation is well calibrated is a topic of debates. These two issues are somehow linked together: in order to effectively measure a recessional velocity due to the Hubble flow, the standard candles must be located far enough so that the Hubble flow makes the dominant contribution to the measurement. Far away located standard candles must be bright enough to be observed from the Earth, and thus the choice is limited. Luminous events are usually quite energetic, and energetic events are scarce. Thus, in a small volume of Universe, there are not many of them available to be studied in details, and properly calibrating their distance-luminosity relation requires the use of intermediate distance calibrators. This is why we speak about a "cosmic ladder", where each step is a distance calibrator that overlap with the previous and following steps, thus giving access to standard candles at an increasingly far distance.

The current state of play uses type-Ia Supernovae (hereafter SNeIa) as the main secondary distance indicator. Kowal (1968) showed first that some Supernovae exhibited a similar absolute luminosity, which was confirmed by Phillips (1993) for a given family, the type-Ia. Building on the discovery of the Universe accelerated expansion by Riess et al. (1998), many surveys aiming at discovering SNeIa were devised. Let us cite for example, among many others, the Canada-France-Hawaii Telescope Supernovae Legacy Survey (SNLS, Astier et al., 2006; Guy et al., 2010) and the SDSS-II Supernovae Survey (Frieman et al., 2008) that have been recently reanalyzed and merged into the Joint Light-curve Analysis sample (JLA, Betoule et al., 2014), or the SH0ES team (Macri and Riess, 2009; Riess et al., 2011, 2016) that uses the Hubble Space Telescope. The intermediate steps, or *rungs*, to calibrate the distance to SNeIa are multiple. Riess et al. (1998, 2016) use for example Cepheids, i.e. stars whose absolute luminosity is related to periodic variations of their luminosity (Leavitt (1908), or in a less distant past Cox (1980); Madore and Freedman (1991)), geometric distance indicators like the distance to stars in the Milky Way and Magelleanic Clouds using parallax measurements, and physical size of stars in binary pairs (Detached Eclipsing Binaries, see Paczynski, 1997). On the other hand, the Carnegie-Chicago Hubble Program (Beaton et al., 2016) aims at using other intermediate calibrators: stars located at the Tip of the Red Giant Branch (TRGB, Lee et al., 1993), i.e. stars at a particular time at the end of their lives where they reach an absolute luminosity sufficiently well known so that they can be used as standard candles (Hatt et al., 2017; Jang et al., 2017). Similarly, RR Lyrae stars are core helium-burning stars that present a relation between the period of their luminosity variations and their absolute luminosity, similarly to Cepheids (see Bono, 2003; Hatt et al., 2017, for a review and a recent example, respectively). Recently, Courtois and Tully (2012); Sorce et al. (2012) proposed to use the empirical correlation between spiral galaxies rotational speed and their intrinsic luminosity - the Tully-Fisher relation (Tully and Fisher, 1977) - as an intermediate calibrator. Finally, it is also possible to use *only* Cepheids as secondary distance indicators, as proposed by Freedman et al. (2012). This has the advantage to remove one step on the distance ladder, and thus reduce the contribution of the step-to-step calibration in the total error budget. In a similar fashion, Chávez et al. (2012) make use solely of the luminosity-velocity dispersion relation of giant extra-galactic regions of ionized hydrogen to infer their distances and thus measure the Hubble constant, building somehow a single-rung ladder. To end that long (yet non-exhaustive) list of possible rungs combination, let us finally mention Dhawan et al. (2017) who propose to revisit the SH0ES team measurements by calibrating the final SNeIa rung not using their luminosity in the visible band but in infrared, where the assumption of SNeIa being standard candles is supposedly more robust.

Cosmic ladders are currently the most precise way to directly measure the Hubble constant in the local Universe. Recently, (Riess et al., 2011; Freedman et al., 2012; Riess et al., 2016) have refined and polished every aspect of the method to reach an impressive precision of a less than five percents. It was, however, at the cost of huge observational efforts, notably from the Hubble Space Telescope. The most difficult part of the analysis being also its core, i.e. the cross-calibration of the various standard candles used in the ladder. It requires to find galaxies where two different types of standard candles can be observed in good conditions. Currently, such galaxies are scarce, making the cross-calibration one of the main source of errors and possible unknown biases, especially on the first local rungs. In addition, the assumptions that the standard candles are indeed standard lie on complex astrophysical mechanisms, that still have to be better explored (Galama et al., 1998; Leibundgut, 2001; Li et al., 2003).

1.3.2 Cosmic Microwave Background radiation

Using observations of the Cosmic Microwave Background to perform precision cosmology became an evidence since the cornerstone discovery of the CMB radiation (Penzias and Wilson, 1965). Sec. 1.1.7 briefly presented the history of the CMB discovery and the pivotal role it plays in modern cosmology. Here, I focus instead on more recent experiments.

The most recent space-based observations of the CMB have been made by the WMAP and Planck satellites. The former is a joint NASA/Princeton University telescope operated between 2001 and 2010, and the latter is an ESA/NASA telescope operated between 2009 and 2013. Like their predecessor COBE, WMAP and Planck observed the temperature fluctuations of the CMB radiation (see Fig. 1.4) and used 2-point correlation function statistics to compare this measurement with the predictions from theoretical models. The CMB is, by definition, a probe of the Universe in the earliest observable stage of its life, around 380’000 years after the Big Bang. Thus, its sensitivity is primarily to the components that were dominant at that epoch. The way pressure waves propagated in the hot soup of baryons, photons and neutrinos that composed the Universe back then, as well as the interactions between these components played an important role in the structures observed in the CMB. The physics of these interactions is not simple (Peebles, 1968; Bond and Efstathiou, 1984; Planck Collaboration et al., 2016c, among many others), hence making the theoretical models quite complex as well. In addition, cleaning the observations to obtain the maps presented in Fig. 1.4 requires to take into account multiple sources of bias, such the presence of the Milky Way (Kogut et al., 1993; Planck Collaboration et al., 2014) or perturbations by foreground emission sources (e.g. Copi et al., 2006; de Oliveira-Costa and Tegmark, 2006; Planck Collaboration et al., 2016b). Overcoming these theoretical and technical difficulties, the WMAP and Planck teams managed to get the most stringent constraint to date on the cosmological parameters (Hinshaw et al., 2013; Planck Collaboration et al., 2016c).

In a complementary fashion, ground-based experiments such as the Atacama Cosmology Telescope (Kosowsky, 2003, hereafter ACT) and South Pole Telescope (Carlstrom et al., 2011, hereafter SPT) carried the same kind of measurements from the ground, trading the thickness of the whole atmosphere for a much larger collector size (their primary mirror are respectively 6- and 10-meters wide, against 1.5-meter for Planck). Both the ACT and SPT, due to their large size, are much more sensitive to small-scale fluctuations in the CMB than their space counterparts (Fowler et al., 2010). They do not cover at all the whole surface of the sky like WMAP or Planck do, but focus instead on smaller, well-selected regions. Doing so, they gain access to scales unreachable by the space instruments, thus making their analysis complementary (e.g. Keisler et al., 2011; van Engelen et al., 2012; Hinshaw et al., 2013).

The great strength of the CMB experiments is that by looking at the Universe at a broad range of physical scales, they are able to constrain all the cosmological parameters at the same time to an unprecedented precision. Yet, since all the parameters are at interplay, one value cannot be individually changed without impacting all the others. This translates into the fact that the CMB experiments do not yield independent determinations of the cosmological parameters, but rather a joint constraint. More importantly in the present context, since the Hubble constant is a *local* parameter whose value relates to the present-time acceleration of the Universe, its determination from CMB measurements requires many other intermediate steps and supplementary assumptions. H_0 is not directly measured from CMB experiments, but rather deduced from a combination of other cosmological parameters. This definitively calls for a comparison with more direct measurements.

It has also to be noted that despite the huge amount of work dedicated to it, the results from Planck are not always in agreement with the ones from WMAP. The two experiments being focused on the same observable, comparing their results allows to cross-check the validity of their respective analysis. In that regard, Copi et al. (2015); Larson et al. (2015), among others, noted that there is a discrepancy between the two experiments. Some authors suggested possible improvements in the Planck analysis pipeline (e.g. Spergel et al., 2015b), that have mostly been taken into account in the most recent Planck data release and analysis. Still, the tension persists (e.g. Addison et al., 2016; Calabrese et al., 2017) and can potentially be the sign that extra considerations and refinements are required in the Planck and/or WMAP pipelines.

1.3.3 Baryon Acoustic Oscillations surveys

Baryon Acoustic Oscillations (hereafter BAO) refer to the fluctuations of density of baryonic matter in the Universe. Simply put, in a nearly homogeneous medium of matter perturbed by pressure waves and subject to gravitational forces, the matter tend to cluster due to the force of gravity. Many clusters of various size and mass are formed in different regions of space, and the mean distance between them depends mainly on the velocity at which the pressure waves travel in the medium. What this implies is that if one understands the underlying physics behind the propagation of the pressure waves at a given epoch, and if one observes at the same epoch how the matter clustered, then one can physically deduce the absolute mean distance between the clusters. This distance can serve as a *standard ruler*, i.e. it assigns a value with physical units to a given distance measurement in the Universe at a given time. Applied to the CMB, the BAO corresponds to the mean size between the over- and under-densities in the CMB map, themselves related to the temperature fluctuations seen in Fig. 1.4. This measurement, called the *sound horizon at recombination*, is a key ingredient used in the CMB analysis to provide constraints on the cosmological parameters and the Hubble constant (Planck Collaboration et al., 2016c).

Following the recombination era - i.e. the period at which the Universe emits the CMB radiation -, two opposite forces are at play on large scales. The force of gravity makes the matter accrete towards the over-densities, thus forming galaxies and clusters of galaxies. The energy released by the Big Bang acts in the opposite direction and inflate the Universe, increasing its size - only marginally helped by the dark energy, whose effect will become dominant much later on in the Universe history. Thus, as the galaxies and clusters become bigger and denser, the mean distance between them, also called the *acoustic scale*, increases over time. Measuring how the acoustic scale evolves allows to constrain the expansion rate of the Universe over time.

Doing so requires to observe how galaxies are arranged in space at different epochs. Since the theoretical prediction of this effect (Peebles and Yu, 1970) and its first observation by COBE in the CMB (Smoot et al., 1992b), many surveys were devised to observe the BAO at various distances and scales. Signals were detected in surveys that measured the spectra of millions of galaxies such the Sloan Digital Sky Survey (SDSS, Eisenstein et al., 2005), the 6-degree Field Galaxy Survey (6dFGS, Beutler et al., 2011), the WiggleZ Survey (Blake et al., 2011), the Galaxy And Mass Assembly survey (GAMA, Liske et al., 2015) or more recently in a branch of the third iteration of SDSS called Baryon Oscillation Spectroscopic Survey (BOSS, Alam et al., 2016). The extremely large sample of galaxies thus observed and the increasing precision of the instruments used, coupled to the relative simplicity of the theory behind this effect has allowed the BAO to become an extremely precise probe of the acoustic scale evolution. In so doing, it directly probes the evolution of the Universe expansion, thus being an excellent proxy to constrain the nature of dark energy and the history expansion of the Universe. Furthermore, upcoming projects like the Taipan galaxy survey (da Cunha et al., 2017), the Wide-Field Infrared Survey Telescope (WFIRST, Spergel et al., 2015a) the Euclid mission (Laureijs et al., 2011), the Large Synoptic Survey Telescope (LSST, LSST Science Collaboration et al., 2009) or the Square Kilometer Array (SKA, Dewdney et al., 2009), to name only a few, will continue the systematic mapping of the Universe in a large range of wavelengths, hence pushing even further the precision to with which BAO will be measured. However, in order to measure H_0, which is an absolute value, the BAO alone are not enough. It needs an absolute calibration of the evolution of the acoustic scale with redshift. Conveniently, it uses the sound horizon at recombination from CMB measurements. Because of this, the BAO cannot be seen as an independent probe of the Hubble constant, but as a complementary method to the CMB that greatly helps refining the value of H_0 (Bennett et al., 2013a; Planck Collaboration et al., 2016c; Alam et al., 2016).

1.3.4 Other techniques

The common point to all the techniques presented so far is that measuring the Hubble constant relates to a measure of an *absolute* distance, somewhere in the Universe. It can be either a radial distance as in the case of the distance ladders, or a transverse distance in the case of the CMB and the BAO. Radial distances are direct applications of the Hubble law, where the distance correlates with the recessional velocity. Transverse distances link the observed angular size of the objects to their physical size, and are thus called angular distances. I present here two techniques other than time-delay cosmography that are also getting closer to the requested precision to play a significant role in the current era of precision cosmology.

The first technique uses water masers. These are seen in some specific types of galaxies with AGN at their core. In the accretion disk of the AGN are maser emitters (like lasers, but emitting in the microwave band), i.e. sources of coherent light emission at a given wavelength. These properties of the light emitted allows to finely pinpoint where the masers are located in the accretion disk, thus making them excellent probes of the structure of the disk, as illustrated by Herrnstein et al. (1999). Under the assumption that the AGN accretion disk follows a Keplerian motion, the maser sources located in the accretion disk exhibit a tangential velocity due to the disk rotation, but also an radial acceleration as they drift towards the center of the AGN over time. Measuring both the tangential velocity and the angular separation between the masers located at the outer edges of the disk, as well as the drift over time towards the center of the AGN of the masers located along our line of sight allows to infer the angular-diameter distance to the galaxy without any direct measurement of the physical size of the accretion disk. The absolute distance, combined to the systemic velocity of the maser can in turn be used to determine the Hubble constant. The current precision of the technique is around 10% for each water maser (e.g. Braatz et al., 2010; Reid et al., 2013; Kuo et al., 2013; Humphreys et al., 2013; Gao et al., 2016). Although the AGN disk model is not yet completely understood, the currently limiting factors are the resolution of the antennas probing the maser emission regions and the number of known galaxies with water maser emission. But if water masers are currently less precise than the other techniques to determine H_0 directly, they still yield an absolute distance to the galaxies hosting the emission that can be used as primary distance indicators for cosmic ladders (e.g. Riess et al., 2016).

The second technique is based on Sunyaev-Zel'dovich effect (hereafter SZ). It predicts that photons emitted by the CMB, when traveling through hot gas around cluster of galaxies have a small chance to encounter high-energy electrons that will give then a boost of energy when they collide, by inverse Compton scattering effect (Sunyaev and Zeldovich, 1972). The same gas through which CMB photons travel also emits in the X-ray band. Both the SZ effect and

X-ray emission depend on the density of electrons in the hot gas integrated along the line-of-sight, but their intensity do not scale similarly with it. Hence, the ratio of the two measured quantities yields a value for the density of electrons in the cluster, which in turn allows to calibrate its luminosity distance and directly measure the Hubble constant from it (Silk and White, 1978; Cavaliere et al., 1979; Hughes and Birkinshaw, 1998), although not with great precision. More recently, (Bonamente et al., 2006) compiled the observational data of 38 SZ clusters into a single measurement of the Hubble constant with a close to 10% precision. The relatively large error on H_0 comes from systematic errors, implying that a better understanding of the intracluster medium is certainly required in order to aim for a better precision.

1.3.5 ...and time-delay cosmography?

Time-delay cosmography is a geometric way to measure the Hubble constant that uses the measurement of the time delays between multiple images of a lensed background source. The general principles have been already presented in Sec. 1.2 as well as some of the related difficulties. Since one of the main topic of my PhD thesis *is* the application of time-delay cosmography, it is not that mere paragraph that will be dedicated to it, but rather the rest of this thesis.

One interesting piece of information that is worth being mentioned before delving into the depths of this technique is that until the last few years, time-delay cosmography was not considered as a mature technique. The controversies on time-delay measurements in the 1980's and 1990's, the approximate lens modeling techniques and all the various resulting H_0 values not agreeing with each other were obviously not helping in that regard. Compared to the huge progresses made by the CMB and BAO measurements and distance ladder technique at the turn of the century, time-delay cosmography was stagnating and not yet ready to enter the era of precision cosmology. It changed, however, a few years later with the start of high-precision and high-cadence monitoring campaigns of lensed quasars (e.g. Fassnacht et al., 2002; Burud et al., 2002) and the emergence of COSMOGRAIL (see Eigenbrod et al., 2005, and Chapter 2), dedicated observations of the quasar host galaxies (Wucknitz et al., 2004; Suyu et al., 2006), detailed modeling from high-quality observations of lens galaxies (Koopmans and Treu, 2003) and a precise study of the impact of other galaxies along the line-of-sight (Keeton and Zabludoff, 2004; Wambsganss et al., 2005; Fassnacht et al., 2006). With a better handle on the sources of error that undermined the field for too long, Suyu et al. (2010, 2014) measured the Hubble constant from two different strong lenses with a $\approx 6\%$ precision on each lens. Finally, Bonvin et al. (2017) combined three strong lenses together and reached a precision on H_0 of 3.8%, making time-delay cosmography once again competitive in the framework of precision cosmology along CMB, BAO and cosmic ladder measurements. Time-delay cosmography is also, along with distance ladder measurements, a *local* probe of cosmic expansion. Being independent from each other, they can thus act as consistency cross-checks, as reviewed in de Grijs et al. (2017).

In the hopefully long history of H_0 estimates from time-delay cosmography, it is certainly not the first time nor the last that authors claim that a few percent precision has been reached. We learned from the past that such claims were, sadly, often not really precise and even less accurate. In what follows, I will do my best to demonstrate why we should now be much more confident in the time-delay cosmography results. Yet, caution should nevertheless still be observed, since it is a well established fact that unexpected sources of error might still be affecting the results. Only time will tell if these unknown unknowns were properly under control, or not.

1.4 Outline

Throughout this manuscript, I detail the work accomplished during my PhD thesis that revolves, without surprises, around time-delay cosmography through a precise measurement of the Hubble constant, nearly independently of any other cosmological parameters. My focus is principally on the results of the COSMOGRAIL collaboration that I have been an integral part of since my first day at work, and of the H0LiCOW collaboration that I joined during the early years of my PhD.

Chapter 2 briefly introduces the relevant mathematical and physical aspects related to strong lensing, then presents the principles and challenges associated to quasar monitoring campaigns. My goal is to explain how, from gravitational lenses discovered somewhere in the sky, high-quality light curves of the lensed images are produced. I illustrate this process by presenting the results obtained on a few selected lenses monitored by COSMOGRAIL.

Chapter 3 carries on with the measurement of time delays from the light curves produced through monitoring campaigns. I detail the operation of `PyCS`, a curve-shifting pipeline I contributed to during my PhD thesis. I notably assessed its abilities in the scope of the Time-Delay Challenge, a competition designed to explore the science capabilities of the upcoming Large Synoptic Survey Telescope. I also illustrate how `PyCS` performs on real data by measuring the time delays on some of the light curves already presented in the previous chapter.

Finally, Chapter 4 presents how the so-obtained time delays can be turned into cosmological constraints. I start by reviewing the present state of research about the Hubble constant. Then, I detail the main improvements performed by H0LiCOW in order to measure a value of H_0 at a 3.8% precision. Building on the results of other state-of-the-art experiments, I show how combining these results hints towards new physics beyond the standard cosmological model, and how such constraints could be tightened further with new results to be acquired in the upcoming years.

2 Monitoring campaigns of lensed quasars

There are three important observables in the determination of the Hubble constant from strong gravitational lensing: i) the measurement of time delays, ii) the high-resolution imaging of the lens galaxy and iii) the study of the lens environment. The main part of my thesis was dedicated to i), i.e. being able to measure as precisely and accurately as possible time delays of lensed quasars. But as the controversy on the time delay of Q0957+561 taught us, nothing precise nor accurate can be done without good quality data. In the present context, the data on which time delays are measured are the light curves of the lensed images of quasars. Designing an efficient monitoring campaign and properly reducing the so-acquired exposures into light curves is thus an essential step, which is the topic of the present chapter.

Although data acquisition and reduction are usually folded together with time-delay measurements into publications, I chose here to clearly separate these in two distinct steps and postpone the time-delay measurement part to chapter 3. I start the present chapter by a brief introduction on the principles and equations of strong lensing, in order to formally explain what the time-delay cosmography is about and why time-delay measurements are so important in the process. Building on these theoretical considerations, I expose a few principles that have their importance when designing monitoring campaigns, which is precisely what is done by the COSMOGRAIL collaboration that I introduce as well. Then, I present in a step-by-step approach how the monitoring data are reduced and transformed into light curves. I illustrate this latter point by presenting the results so obtained from a selection of particularly interesting COSMOGRAIL targets. I conclude this chapter by presenting the principles and first results of an extension of the COSMOGRAIL project developed during the second half of my PhD thesis that aims at speeding-up the data acquisition process through high-cadence monitoring.

2.1 The equations of time-delay cosmography

The equations linking observations to time-delay cosmography are derived from the principles of General Relativity. They were first theorized by Refsdal (1964b,a) and notably rewritten in a new formalism by Schneider (1985) that is still widely used today. Since then, they have been reproduced countless times, more or less successfully. In this section, I do not aim at covering exhaustively the mathematical and physical principles of strong lensing and time-delay cosmography. I rather present an excerpt of what I consider to be the minimal set of equations necessary to understand *why* we measure time delays, and why it is so important in time-delay cosmography. I also present how this mathematical formalism motivates our choice of observables, i.e. which lensed quasars are the most suitable for time-delay cosmography. The formalism presented has been distilled from the very pedagogical introduction presented in Schneider (2006). Let us note that all along this section, we assume that the strong lens systems we are interested in lie in an otherwise homogeneous universe, described with an Friedman-Lemaître-Robertson-Walker (or FLRW) metric. A metric is an exact solution of the Einstein's field equations of General Relativity. The FLRW metric assumes large-scale homogeneity and isotropic expansion (or contraction).

2.1.1 The lensing equation

Let us start by drafting a sketch of what our typical galaxy-quasar-lens configuration looks like. This is presented in Fig. 2.1.

The observer & lens, observer & source and lens & source are separated by the *angular diameter distances* D_{d}, D_{s} and D_{ds}, respectively, whose proper definition is postponed to later on in this manuscript. The projected angle between the lens and the source is labeled $\boldsymbol{\beta}$ and corresponds to a projected distance $\boldsymbol{\eta}$ in the source plane. A light ray emitted by the source reaching the lens plane at a distance $\boldsymbol{\xi}$ or angle $\boldsymbol{\theta}$ is deflected by an angle $\hat{\boldsymbol{\alpha}}$. Let us note that in reality, the bending of the light ray occurs smoothly all along its passing through the gravitational field of the lens galaxy. However, the distances D_{d} and D_{ds} are much larger than the physical size of the lens mass distribution and associated gravitational field. We speak in this case of a *geometrically thin lens*, for which the approximation of a mass distribution in a plane and a single deflection angle $\hat{\boldsymbol{\alpha}}$ holds very well. Its amplitude $\hat{\alpha} = ||\hat{\boldsymbol{\alpha}}||$ is predicted by General Relativity, in the case where the *impact parameter* $\xi = ||\boldsymbol{\xi}||$ is much larger than the Schwarzschild radius $r_s = 2GM/c^2$, i.e. the radius of a sphere of mass M so that the escape velocity at its edge is c.:

$$\hat{\alpha} = \frac{4GM}{c^2 \xi}. \tag{2.1}$$

Eq. 2.1 assumes that all the mass of the lens is spatially concentrated into a point. In reality, this is of course not the case; the lens mass distribution is spatially extended in three dimensions. Each mass element $\delta m'$ of the lens total mass distribution located at the coordinate (ξ'_1, ξ'_2, ξ'_3) deflects the passing light rays. We can show that in the thin lens approximation, the deflection

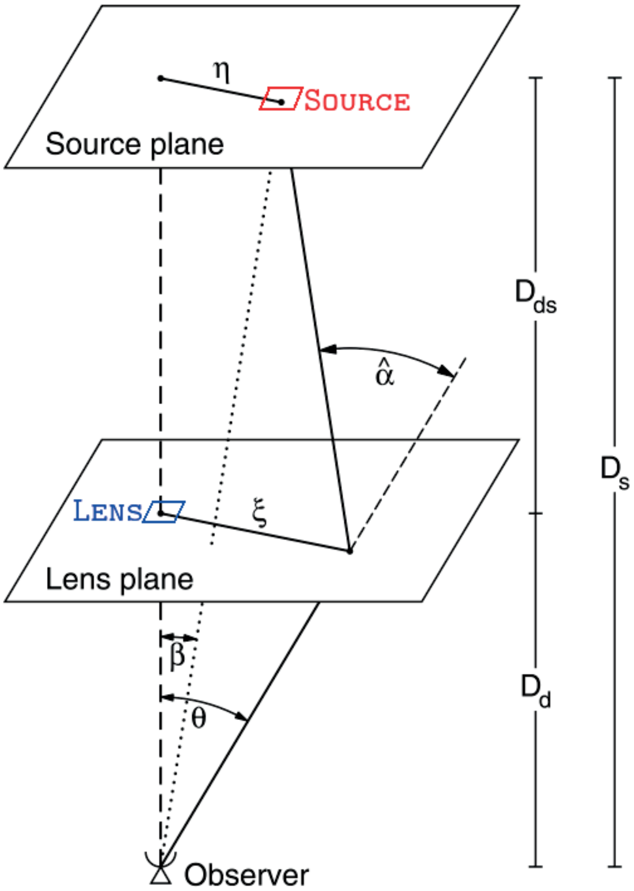


Figure 2.1: Sketch of a gravitational lens system. The dashed line connects the observer with the lens galaxy (blue square), whereas the dotted line connects the observer with the source (red square). The solid line represents the trajectory of a light ray emitted by the source and reaching the observer, bent by an angle $\hat{\alpha}$ when passing by the lens. Adapted from Schneider (2006).

angle can be expressed as

$$\hat{\alpha}(\boldsymbol{\xi}) = \frac{4G}{c^2}\int d^2\boldsymbol{\xi}'\Sigma(\boldsymbol{\xi}')\frac{\boldsymbol{\xi}-\boldsymbol{\xi}'}{|\boldsymbol{\xi}-\boldsymbol{\xi}'|^2}, \qquad \Sigma(\boldsymbol{\xi}') = \int d\xi_3'\rho(\xi_1',\xi_2',\xi_3'), \tag{2.2}$$

where $\Sigma(\boldsymbol{\xi}')$ is the *projected surface mass density* at the two-dimensional coordinate $\boldsymbol{\xi}'$, and $d^2\boldsymbol{\xi}'$ is the 2-dimensional distance element connecting the center of the lens mass distribution to the orthogonal projection on the lens plane of each mass element $\delta m'$. Basically, all the mass of the lens is projected into the lens plane and the deflection angle is computed by integrating the effect of each projected mass element.

Eq. 2.2 expresses the amplitude of the deflection angle of a light ray $\hat{\alpha}(\boldsymbol{\xi})$ in function of the physical distance $\boldsymbol{\xi}$ at which it crosses the lens plane. It is, however, much more convenient for the observers to express it relatively to its angular equivalent $\boldsymbol{\theta}$ that can be directly measured. This can be done under the approximation that all the angles considered in this geometric problem are very small, which once again holds very well considering the astronomical distances involved in real-life scenarios:

$$\boldsymbol{\eta} = D_{\mathrm{s}}\boldsymbol{\beta}, \qquad \boldsymbol{\xi} = D_{\mathrm{d}}\boldsymbol{\theta}, \qquad \boldsymbol{\eta} = \frac{D_{\mathrm{s}}}{D_{\mathrm{d}}}\boldsymbol{\xi} - D_{\mathrm{ds}}\hat{\boldsymbol{\alpha}}(\boldsymbol{\theta}). \tag{2.3}$$

Folding the terms of Eq. 2.3 together, we obtain the so-called *lens equation*:

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \frac{D_{\mathrm{ds}}}{D_{\mathrm{s}}}\hat{\boldsymbol{\alpha}}(D_{\mathrm{d}}\boldsymbol{\theta}) \equiv \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta}), \tag{2.4}$$

where we introduced the *scaled deflection angle* $\boldsymbol{\alpha}(\boldsymbol{\theta})$. The amplitude of the latter can thus be expressed, using Eq. 2.2 as follows:

$$\alpha(\boldsymbol{\theta}) = \frac{1}{\pi}\int d^2\boldsymbol{\theta}'\kappa(\boldsymbol{\theta}')\frac{\boldsymbol{\theta} - \boldsymbol{\theta}'}{|\boldsymbol{\theta} - \boldsymbol{\theta}'|^2}, \qquad \kappa(\boldsymbol{\theta}) = \frac{\Sigma(D_{\mathrm{d}}\boldsymbol{\theta})}{\Sigma_{\mathrm{crit}}} \qquad \Sigma_{\mathrm{crit}} = \frac{c^2}{4\pi G}\frac{D_{\mathrm{s}}}{D_{\mathrm{d}}D_{\mathrm{ds}}}, \tag{2.5}$$

where $\kappa(\boldsymbol{\theta})$ is called the *convergence* and Σ_{crit} the *critical mass density*. The integration is made this time over the angles $\boldsymbol{\theta}'$, which can be directly measured by the observer. At this stage, we can already see that the Eqs. 2.4 and 2.5 provide a mapping between the observed lensed images position $\boldsymbol{\theta}_i$ and the source position $\boldsymbol{\beta}$. In a system where multiple lensed images are seen, these relations must yield a similar $\boldsymbol{\beta}$ for each $\boldsymbol{\theta}_i$, thus constraining the projected mass distribution Σ at the images position $\boldsymbol{\theta}_i$.

Before going further, it is interesting to consider the case where the source, lens and observer are perfectly aligned. In such a case, we can deduce from symmetrical considerations that an "infinite" number of merging images of the source are visible all around the lens, all seen at the same distance of the lens; they appear to us under the form of a *ring*. Since the alignment is perfect, $\boldsymbol{\beta} = 0$. The lens equation 2.4 thus yields $\boldsymbol{\theta} = \boldsymbol{\alpha}(\boldsymbol{\theta})$. Folding it into Eq. 2.1, we can express the angular size of the ring as:

$$\theta_E = \sqrt{\frac{4GM}{c^2}\frac{D_{\mathrm{ds}}}{D_{\mathrm{s}}D_{\mathrm{d}}}}. \tag{2.6}$$

θ_E is called the *Einstein radius*. In real-life cases where the lensed system configuration is very close to a perfect alignment, one might observe both lensed images of the quasar at positions $\boldsymbol{\theta}_i$ and part of an Einstein ring of radius θ_E. Measuring the latter gives us access to the total mass M "enclosed" into it.

2.1.2 The time-delay distance

The writing of Eq. 2.4 using the scaled deflection angle $\boldsymbol{\alpha}(\boldsymbol{\theta})$ and its expression in Eq. 2.5 are not completely fortuitous. Let us now introduce two interesting quantities. The first one is the *deflection potential* ψ:

$$\psi(\boldsymbol{\theta}) = \frac{1}{\pi}\int d^2\boldsymbol{\theta}'\kappa(\boldsymbol{\theta}')\,ln|\boldsymbol{\theta}-\boldsymbol{\theta}'| \tag{2.7}$$

This quantity is called a potential by analogy with classical mechanics; if the scaled deflection angle plays the role of the acceleration and the convergence the role of the mass, then the relation

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \nabla\psi(\boldsymbol{\theta}), \tag{2.8}$$

is equivalent to the well-known $\nabla\psi = m\boldsymbol{a}$ for a test mass of $m = 1$. The deflection potential is also often called the *gravitational potential* of the lens.

The second quantity is called the *Fermat potential* τ:

$$\tau(\boldsymbol{\theta},\boldsymbol{\beta}) = \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\beta})^2 - \psi(\boldsymbol{\theta}). \tag{2.9}$$

That quantity has two very interesting properties (demonstrated in Schneider (1985) but not reproduced here). First, writing $\nabla\tau(\boldsymbol{\theta},\boldsymbol{\beta}) = 0$ is equivalent to the lens equation 2.4. Thus, providing $\boldsymbol{\beta}$ and ψ are known, the position of the lensed images can be predicted to appear at the extrema and saddle points of the Fermat potential. Second, $\tau(\boldsymbol{\theta},\boldsymbol{\beta})$ is, up to an affine transformation, the expression of the travel time of the light ray from the source to the observer. Thus, lensed images form at the stationary points of the arrival time surface, notably where the arrival time is the shortest which explain the name given to $\tau(\boldsymbol{\theta},\boldsymbol{\beta})$ by analogy with the Fermat principle[1].

From these mathematical developments, one can express more precisely the *excess of time*

---

[1] The Fermat principle states that the path followed by a light ray between two points is the one that minimizes the travel time.

delay t of a lensed light ray respectively to the case of no lensing, i.e. if the light ray could reach the observer in a straight line. That excess of time delay can be understood as the combination of the *geometrical time delay* due to the path traveled by the light ray being longer than a straight line, and the *Shapiro time delay* due to the light ray traveling through the gravitational potential of the lens and thus moving slower as seen from the outside of this potential:

$$t = \frac{1}{c}\frac{D_{\mathrm{d}}D_{\mathrm{s}}}{D_{\mathrm{ds}}}(1+z_{\mathrm{d}})\left[\frac{(\boldsymbol{\theta}-\boldsymbol{\beta})^2}{2} - \psi(\boldsymbol{\theta})\right], \tag{2.10}$$

where z_{d} is the redshift of the lens, and where we find in brackets the expression of the Fermat potential $\tau(\boldsymbol{\theta},\boldsymbol{\beta})$. Obviously, it is impossible to measure an excess of time delay from a single lensed image. However, the *difference of excess of time delays*, simply abridged to *time delay* Δt_{ij} between two lensed images i and j can be written like this:

$$\Delta t_{ij} = \frac{D_{\Delta t}}{c}\left[\frac{(\boldsymbol{\theta}_i-\boldsymbol{\beta})^2}{2} - \psi(\boldsymbol{\theta}_i) - \frac{(\boldsymbol{\theta}_j-\boldsymbol{\beta})^2}{2} + \psi(\boldsymbol{\theta}_j)\right], \qquad D_{\Delta t} = \frac{D_{\mathrm{d}}D_{\mathrm{s}}}{D_{\mathrm{ds}}}(1+z_{\mathrm{d}}). \tag{2.11}$$

$D_{\Delta t}$ is called the *time-delay distance*, and is the quantity related to the cosmological parameters through the ratio of angular diameter distances. Specifically, $D_{\Delta t} \propto H_0^{-1}$ as demonstrated in Sec. 4.2.4.

Equation 2.11 can be clearly separated in three parts. To the left stands the time delay which can be directly measured. To the right stand the time-delay distance into which the cosmological parameters are folded, and in brackets are the angular coordinates of the lensed images as well as the gravitational lens potential. From the knowledge of two of these parts, the third one can be determined through the equation expressed above. Since the end goal of time-delay cosmography is to measure H_0 which appears in the time-delay distance expression, it becomes clear that a precise and accurate measurement of the time delay is essential.

So far, we considered the source as emitting single light rays in various direction from the same emission point, i.e. as a *point source*. Although in practice the apparent size of the source is extremely small, it is not a real point source. Even if the apparent size of the images of the source are smaller than the pixel size of the CCD detector, they nevertheless are the sum of many different light rays emitted by the source. One very interesting property of the lensing equations is that they conserve the *surface brightness*. It means that if the images are seen bigger than in the case of no lensing, they also appear brighter. The brightness ratio of a lensed image and its non-lensed hypothetical counterpart is called the *magnification* $\mu(\boldsymbol{\theta})$. In the formalism presented above, one can show that the magnification is related to the lensing potential. Light rays coming from a slightly different position in the source plane and crossing the lens plane at a slightly different angle will be deflected slightly differently, yet might still focus towards the observed. Let us introduce the *shear* $\gamma(\boldsymbol{\theta})$ as follows:

$$\gamma(\boldsymbol{\theta}) = \gamma_1(\boldsymbol{\theta}) + i\gamma_2(\boldsymbol{\theta}), \qquad \gamma_1(\boldsymbol{\theta}) = \frac{1}{2}(\psi_{11}(\boldsymbol{\theta}) + \psi_{22}(\boldsymbol{\theta})), \qquad \gamma_2(\boldsymbol{\theta}) = \psi_{12}(\boldsymbol{\theta}), \qquad \psi_{\mathrm{ij}}(\boldsymbol{\theta}) = \frac{d^2\psi\boldsymbol{\theta}}{d\theta_i d\theta_j}. \tag{2.12}$$

The shear, so defined completes the convergence $\kappa(\boldsymbol{\theta})$ defined in Eq.2.5 in the expression of how a lensed image is affected by the lens potential. Roughly speaking, the convergence relates to the apparent size of the lensed image whereas the shear relates the the image's skewness. Thus, both can be combined to yield the magnification $\mu(\boldsymbol{\theta})$:

$$\mu(\boldsymbol{\theta}) = \frac{1}{(1-\kappa(\boldsymbol{\theta}))^2 - |\gamma(\boldsymbol{\theta})|^2}. \tag{2.13}$$

Generally, the absolute magnification cannot be measured, since the source absolute brightness is unknown. The relative magnification of the images, however, is accessible to the observer, and provides an additional constraint that must be reproduced by the lens potential $\psi(\boldsymbol{\theta_i})$ in Eq. 2.11.

To conclude, let us recall that no definition were given yet for the angular diameter distances D_{d}, D_{s} and D_{ds}. In the FLRW Universe that is not stationary, computing these distances is not straightforward and the apparently simple relation $D_{\mathrm{d}} + D_{\mathrm{s}} = D_{\mathrm{ds}}$ does *not* hold true ! However, since a proper definition of these distances is not essential for this chapter, it is postponed to Sec. 4.2.4. Similarly, the formalism described here consider only one single isolated lens; to properly take into account the lens environment, Eq. 2.11 must be adapted. This will be considered later on in Secs. 4.2.2 and 4.2.3.

2.1.3 Choosing the best quasars to observe

The time-delay distance in Eq. 2.11 being at the center of time-delay cosmography, it should naturally drive our choice of quasars on which the cosmography can be carried out. Indeed, since $\Delta t_{ij} \propto D_{\Delta t} \propto H_0^{-1}$, a given relative precision on the time-delay measurement corresponds to an absolute precision on H_0. Precise measurements of H_0 can thus be more easily achieved on systems with longer time delays, since reaching a smaller relative precision is easier. Similarly, the precision of the *astrometry*, i.e. the determination of the images angular coordinates $\boldsymbol{\theta}_i$ and a precise knowledge of the lens potential ψ are directly proportional to the precision on H_0. Hence, an isolated lens galaxy exhibiting a large angular separation between the lensed images of the background source is likely to yield better results. The good point is that in a given lens configuration, the time delays are usually longer when the images separation are wider.

In practice though, things are not that simple. First, lensed quasars are rare. Only a few hundreds of background objects lensed by single foreground galaxies have been discovered to date, and not all of them are quasars. The current number of discovered quasar-galaxy lenses is around a hundred, and although that number should be significantly pumped up in a near future (see e.g. Oguri and Marshall, 2010; Collett, 2015), not all of these systems are or will be well suited to be monitored; the images might be too faint to be observed with a sufficient precision, they might be outshined by the lens galaxy or they could be too close to each other to be properly distinguished. Of course, the constraints on the minimal luminosity and angular separation of the lensed images depend on the telescope chosen for the monitoring. Thus, the choice of targets to monitor must be adapted to the choice of instrument carrying the monitoring.

Another important factor to consider is the lens configuration. Depending on the alignment between the observer, the lens galaxy and the source, the number of lensed images can vary. Theoretical computations predict an odd number of lensed images (Burke, 1981). This excludes the direct image of the source, since it is usually masked by the lens or simply too faint to be observed directly. The observed lens systems, however, often display an even number of images since the extra one is predicted to be strongly demagnified (Schneider, 1985), and thus way too faint to be observed with the current telescopes sensitivity. This signifies that the lens systems on which time-delay cosmography is currently applicable with sufficient precision display either two or four images. We thus refer to them as *double* or *quad*, to keep it short.

Both doubles and quads have their own pros and cons. Doubles appear when the observer-lens-source alignment is weaker, thus often yielding longer time delays. The smaller number of lensed images makes that the photometric measurement of their luminosity variations is easier, in the sense that the various components have an higher angular separation and are more easily distinguished from each other. From quads, however, up to six time delays can be measured, yet only three of them being truly independent. Combined with the fact that the four images can appear at a different angular distance from the center of the lens galaxy, the number of constraints usable for the lens modeling is more than twice greater. In addition, since quads appear when the observer-lens-source alignment is closer to perfect, they are more prone to exhibit - at least part of - an Einstein ring whose radius is directly linked to the mass of the lens galaxy (see Eq. 2.6). A visible Einstein ring is in fact a very desirable property of a lens system, since it brings an extremely valuable constraint on the gravitational potential of the lens. However, quads are more difficult to model and monitor than doubles due to the usually smaller angular separation between the lensed images themselves as well as the lensed images and the lens galaxy. In addition, the rareness of suitable quads - only roughly one sixth of lensed quasars are quads, according to Oguri and Marshall (2010) - makes us believe that we are using so far only the brightest of them, which could introduce a bias on our general assumptions on quads modeling that propagates on the H_0 determination (Collett and Cunnington, 2016).

Figure 2.2 presents colored images of four lensed quasars (Suyu et al., 2016), three quads and one double all seen in the visible by the *Hubble Space Telescope*. The galaxy is always at the center, and the lensed images appear around it. Quads can appear in three different configurations: RXJ1131-1231 (top left), with three images close to each other and a fourth, fainter one on the other side of the lens galaxy is in a *cusp* configuration. HE0435-1223 (top right), where the four images are nearly symmetrically distributed around the lens galaxy is in a *symmetric* configuration. Finally WFI2033-4723 (bottom left), with two images on one side of the lens galaxy really close to each other, and two other images on the other side with a wider separation, is in a *fold* configuration. For the three quads, we clearly see the Einstein ring passing through the lensed images, that is not seen in the double HE1104-1805. Suyu et al. (2016) shows however that there is an Einstein ring around HE1104-1805, but much fainter than those observed in the quads. In the case of RXJ1131-1231 we can even see different parts of the source host galaxy appearing in the Einstein ring. The three quad configurations also have an impact on the expected time delays. For example, symmetric configurations are expected to yield shorter time-delay measurements since the configuration is closer to symmetry; cusp configurations have longer time delay between the faintest image and the three others.

All these theoretical considerations are worth being mentioned, yet it is important to keep in mind that at the time this manuscript is written no more than a few dozens of quasars are suitable to be monitored, doubles and quads alike. Thus, we cannot currently afford being too picky about our choice of targets. This is, however, expected to change in a near future, with upcoming large sky surveys that will cover the sky in an unprecedented way and discover a great amount of new strongly lensed quasars.

2.2 COSMOGRAIL

COSMOGRAIL, standing out for the COSmological MOnitoring of GRAvItational Lenses is a collaboration initialized and led by the Laboratory of Astrophysics of EPFL (LASTRO). It started in 2004, under the leadership of Prof. George Meylan and now of Prof. Frédéric Courbin. Its goal is to measure the time delays of lensed quasars with an unprecedented accuracy using small and medium-size telescopes located all around the world, and use these measurements to determine the Hubble constant at a few percents precision. Since its start, it produced so far 16 published papers. Some publications focus on the technical aspects of the method such as the optimal sampling cadence (Eigenbrod et al., 2005) or the time-delay measurement techniques (Tewes et al., 2013a; Bonvin et al., 2016). Other papers address the properties of the lens galaxies in the COSMOGRAIL lensed quasar sample, by measuring their redshift (Eigenbrod et al., 2006a,b, 2007) or modeling their mass distribution (Saha et al., 2006; Chantry et al., 2010; Sluse et al., 2012a). And obviously, a third group of paper focus on time-delay measurements themselves (Vuissoz et al., 2007; Courbin et al., 2011; Eulaers et al., 2013; Tewes et al., 2013b), sometimes accompanied with an associated value of the Hubble constant (Vuissoz et al., 2008; Rathna Kumar et al., 2013). In addition, other publications directly use

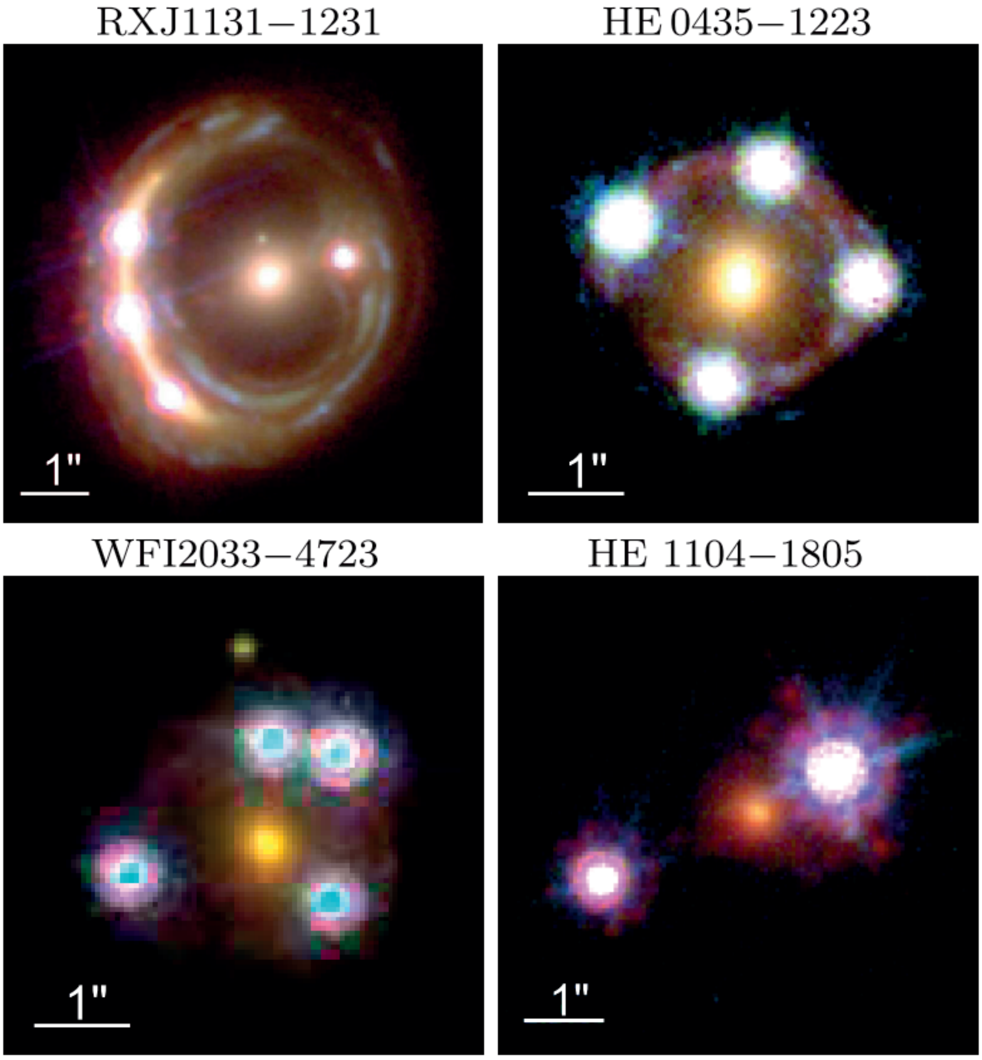


Figure 2.2: Illustration of the various possible configurations of lens systems that can be observed. *T*op-left: quad in a *cusp* configuration. *T*op-right: quad in a *symmetric* configuration. *B*ottom-left: quad in a *fold* configuration. *B*ottom-right: double. All the images are a stack of Hubble Space Telescope images in visible bands. Adapted from Suyu et al. (2016).

the COSMOGRAIL data to determine the Hubble constant (e.g. Suyu et al., 2014; Birrer et al., 2016). Among them are namely the results of the H0LiCOW collaboration (Suyu et al., 2016; Bonvin et al., 2017) built around COSMOGRAIL and that is described in much more details in Sec. 4.2.

2.2.1 Monitoring cadence

From 2004 to 2011, five telescopes participated in COSMOGRAIL: the Himalayan Chandra Telescope in India, the Mercator telescope and Liverpool Robotic Telescope at La Palma in

Spain, the AZT-22 telescope on Mt. Maidanak in Uzbekistan and the Leonhard Euler Swiss telescope (hereafter Euler) at La Silla in Chile. Since 2012, only Euler carried on with the monitoring; it was the only facility that was able to keep up with the desired monitoring cadence over long periods of time, and the installation of the new camera EulerCAM (hereafter ECAM) in 2011 made it significantly more precise than the other instruments used since then. The monitoring cadence was chosen as a function of the achievable precision of the photometric measurements from the various instruments involved in the observations. From the firsts long-term monitoring campaigns in the 1980's and 1990's, the quasars' luminosity variations have been exhibiting both long-term variations of large amplitude and shorter variations of smaller amplitude (Hook et al., 1994). According to this fact, Eigenbrod et al. (2005) showed that observations up to twice a week were the optimal choice to precisely measure a time delay after only a few years of monitoring. It turned out that for a variety of reasons (notably the presence of microlensing), the required baseline to obtain an accurate measurement ended up being closer to a decade than a single year, although the original bi-weekly cadence was the most appropriate choice. As of today, that cadence has been slightly reduced notably due to the increasing number of discovered quasars and the limited time available at Euler for monitoring, yet each target is still observed at least once per week.

For completeness, let us mention here that in a complementary fashion, an extension to the original program has been developed in order to measure time delays in a much shorter period of time. Using first the Very Large Telescope (hereafter VLT) at Paranal in Chile for a 6-months pilot monitoring campaign in 2014/2015 and then the 2.2m ESO/MPG telescope (hereafter 2m2) at La Silla since fall 2016, this new observing strategy increases the monitoring cadence to one observation per night, and aims for a photometric precision close to the mili-magnitude. So doing, we expect to be able to measure precise time-delays in less than a year. This new strategy is developed further in Sec. 2.5.

2.2.2 Choice of targets

As explained above, the scarcity of discovered lensed quasars implied that the choice of targets to monitor has been driven mainly by telescope specifications: each lensed quasar that is bright enough is a potential target, providing that it is visible from a long enough period of time each year from the place the monitoring telescopes are located. Over the years following the start of COSMOGRAIL, new lensed quasars were discovered thus incrementing the pool of targets, whereas older targets were abandoned because of the absence of good results (e.g. no observed luminosity variations or variations dominated by microlensing). On January 1st 2017, the number of lensed quasars still being monitored by COSMOGRAIL reached the number of 20. In total, 44 quasars were observed by COSMOGRAIL since its beginning in 2004. Table 2.1 presents the status of the targets being currently monitored. All of them are located either in the South galactic hemisphere or close to the galactic equator since they are observed from the southern Earth hemisphere. As predicted by (Oguri and Marshall, 2010), quads are less common than doubles, and this is well represented in this table where only 1/4th of the sample are quads.

Table 2.1: List of all the COSMOGRAIL lenses currently being monitored at the Swiss 1.2m Leonhard Euler Telescope, along with the number of observing nights dedicated to each lens (usually 1 night ≃ 30 minutes of observation) and the number of years elapsed since the first observation. The lens configuration is also reported for each lens, where *fold, cusp* and *symmetric* refer to quads configuration. The (?) associated to DES J0408-5354, a recently discovered quasar (Lin et al., 2017; Agnello et al., 2017) refers to the fact that it is currently unclear which configuration is exhibited by the system.

Name	# observing nights	# years of monitoring	configuration
DES J0408-5354	39	1.2	fold (?)
DES2327-5248	27	1.1	double
HE0047-1756	682	12.6	double
HE0435-1223	878	13.2	symmetric
HE1104-1805	114	3.2	double
HE2149-274	565	12.6	double
HS0818+1227	408	12.4	double
J0158-4325	503	12.3	double
J0246-0825	355	10.2	double
J0832+0404	209	6.9	double
J1226-0006	495	12.2	double
J1335+0118	553	13.0	double
J1349+1227	135	6.9	double
J1405+0959	45	3.0	double
J1455+1447	128	6.9	double
J1515+1511	48	3.0	double
J1620+1203	186	6.5	double
RXJ1131-1231	740	13.1	cusp
WFI2026-4536	595	12.6	fold
WFI2033-4723	664	12.6	fold

For a similar number of years since the start of the monitoring, the number of observing epochs can vary. For example, after 13 years of monitoring, HE0435-1223 and RXJ1131-1231 have been observed twice more frequently than J1226-0006 or J1335+0118. This is explained by the fact that some quasars are more interesting than others, in the sense that they either rapidly exhibited evident signs of strong variability, or that the lens parameters (redshift, mass, etc...) were available soon after the start of the monitoring, thus making the system suitable for time-delay cosmography. As stated above, quads have naturally more observational constraints available, which makes them more interesting than doubles for cosmology, and this stands out in the table where all the quads were observed more frequently than doubles. Note that ideally, all the targets would have been observed with the same frequency, but the limited amount of observing time forced us to proceed otherwise.

2.3 Data reduction

Once the right cadence has been chosen and the targets regularly observed, we still need to extract the relevant scientific information from the observations. In the present case, that corresponds to a precise photometric measurement of the luminous flux of the various lensed images, extracted from the raw exposures. Since what we want to measure are the night-to-night relative variations between the lensed images, the measured flux do not need to be absolutely calibrated. However, we need to take care of removing all the systematic effects that might affect the relative calibration. The processing of the exposures is done through a largely automated pipeline described in this section.

2.3.1 The COSMOULINE pipeline

The origins of COSMOULINE date back to the PhD thesis of Malte Tewes (Tewes et al., 2013c) who developed the pipeline, and to which I added both scientific and technical improvements. In the following, I briefly review the main steps of the process, that I illustrate through their application on real data, notably from the monitoring of HE0435-1223 that were used to measure the time delays presented in Bonvin et al. (2017). Let us note that the whole pipeline is based only on a few softwares. Apart from `python`, we use *Sextractor* (Bertin and Arnouts, 1996) to handle the identification and automatic aperture photometry of the sources in the field, the *Astropy* python package (Astropy Collaboration et al., 2013) to handle the .fits images operation and the *MCS deconvolution* package (Magain et al., 1998; Cantale et al., 2016) to generate the point-spread function on each exposure and deconvolve the lensed images. Although we do not plan to intensively advertise for the existence of COSMOULINE, we made the code publicly available on the COSMOGRAIL webpage[2]. The main steps of the pipeline are as follows.

1. The first obvious step, common to every photometric reduction procedure is to remove the *bias* level from all the exposures, and correct the science exposures from the pixel-to-pixel CCD efficiency variations by taking a *flat-field*, i.e. an exposure of an homogeneously illuminated surface. By ensuring that enough bias exposures and flat-fields were taken in complement to each science exposures, combination of biases called *masterbiases* are constructed, then flat-fields and science exposures are corrected from the bias, masterflat-fields are constructed from multiple flat-fields, and science exposures are finally corrected from the pixels efficiency variations. Providing the acquisition camera is sufficiently cooled - typically with liquid nitrogen -, the "dark" current, i.e. the additional noise generated by the instrument electronic components accumulated during the science exposure can be neglected. This first standard cleaning procedure leaves us with science exposures blurred by the atmosphere and pixelated by the CCD, that contain the flux from the observed objects (including the lensed images

---

[2] http://cosmograil.org/software

of the quasar, obviously), still affected by a number of potential sources of noise that need to be removed.

2. The second step consists of removing these extra sources of noise, which can be separated in two sub-steps. First, there is a large-scale pattern commonly called the *sky* affecting all exposures. Typically, this smooth, large-scale contribution comes from surrounding sources of light or from the diffusion of the moon light in the Earth's atmosphere. Obviously, the sky pattern changes from exposure to exposure, and needs to be modeled individually on each exposure. Second, there is an chromatic effect linked to the variations of thickness of the CCD that can occasionally produce a *fringing* pattern on the CCD. The stability of the pattern depends on the optical elements of the telescope, but is generally stable from night to night. Yet, its intensity varies and the correction must be individually adapted to each exposure.

   Both the sky and the fringing patterns are additive sources of noise, so once computed they can be simply subtracted from the science exposures. The sky pattern is created using *Sextractor* by computing the median count per pixel in large overlapping portions of the exposures. The fringing pattern is computed on already sky-subtracted images, by taking the median of a stack of many sigma-clipped exposures of different portions of the sky in order to remove the contribution from the stars. Among the telescope used for the monitoring, only the 2m2 occasionally suffered from fringing effect and thus was the only one that needed the appropriate correction. Figure 2.3 illustrates the sky subtraction process on an exposure of HE0435-1223 taken at Euler, and the fringing subtraction process on an exposure of HE0047-1756 taken at the 2m2.

3. The third step consists of aligning all the exposures, and select the best stars around the lens galaxy to estimate the Point-Spread Function (PSF) on each exposure. The PSF represents a model of the diffraction of a theoretically point source of light, due to the passing of the light rays through the atmosphere and their interaction with the optical elements of the telescope. Stars located far enough from the observer have an angular size much smaller than the pixel size of the detector. In ideal conditions, they should thus appear as a "point", i.e. only visible on one pixel of the camera. In practice, the images of such stars are broadened by the effects mentioned above, and appear as luminous blobs on the detector. For ground-based telescopes, the diffraction from the atmosphere is the main contributor to the broadening of the PSF. On large-size ground telescopes like the VLT or Keck, the broadening can be partially mitigated using Adaptive Optics (Beckers, 1993), i.e. an instant deformation of the telescope mirrors in order to correct from the deformations in the incoming wavefront. Other non-dominant effects are also contributing to the PSF. For example, the color of the sources chosen to model the PSF can change its width. The position of the sources on the detector affects as well their individual shape, usually distorting more and more the PSF when the sources are located farther away from the center of the CCD.

   To deal with these accumulated effects as efficiently as possible, the PSF is constructed using multiple stars in the field of view of each target, if possible from the same color

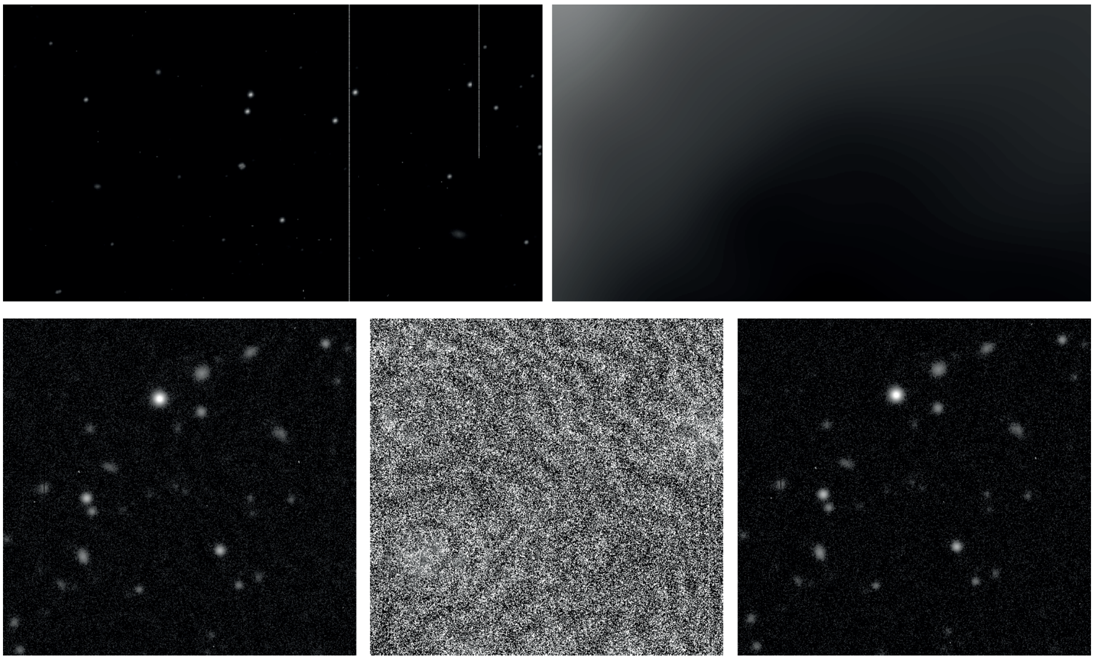


Figure 2.3: The top row illustrates the sky subtraction process on a cut of a single 6-minutes exposure of HE0435-1223 taken at Euler. The left panel represents a region before the sky subtraction gets applied, and the right panel present the modeling of the sky on the same region. The noise from the sky, although usually not visible by eye on the uncorrected exposures, can have a dramatic effect on the precision of the final measurements if not properly accounted for. The bottom row illustrates the fringes subtraction process on a cut of a single 6-minutes exposure of HE0047-1756 taken at the 2m2. The leftmost panel presents the already sky-subtracted region, the middle panel presents the modeling of the fringes on the same region (with an enhanced contrast for visual purposes), and the right panel presents the region with the fringes removed. The difference between the left and right panel are clearly visible by eye.

and luminosity than the lensed images. The PSF fitting is done in two steps. First, a common analytical profile, such as a Moffat (A Moffat profile, compared to a Gaussian is sharper at its center and has wider tails) is iteratively fitted to all sources. Second, a regularized fine-pixel array is added to the analytical profile in order to improve the fit. During this procedure, the sampling of the PSF is forced to be twice the one of the original pixelated images of the sources, and the modeling of the PSF is such that a "perfect" point source, once corrected from the PSF - i.e. *deconvolved* - is represented by a circular Gaussian function with a Full-Width at Half Maximum (FWHM) of 2 pixels. That method has proven to be very efficient at reducing the noise in the deconvolved images - a very important point as illustrated in the following step of the pipeline.

Figure 2.4 displays the field-of-view of the lensed quasar HE0435-1223 seen through the ECAM instrument mounted on the Euler telescope. The quasar is in the square at the center, and the stars chosen for the PSF modeling are labeled in red.

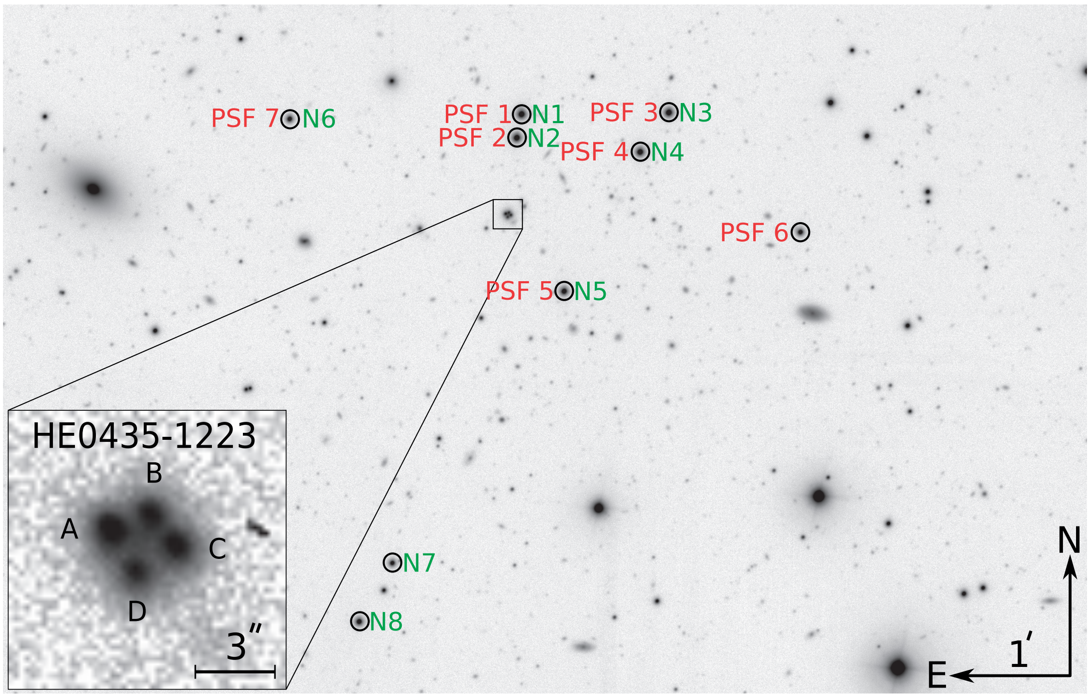


Figure 2.4: Part of the field-of-view of HE0435-1223 seen through the Euler telescope. This image is a combination of 100 exposures of 360 seconds each, for a total exposure time of ∼10 hours. The stars used to build a PSF model are circled in red and labeled PSF1 to PSF7, and the stars used for the photometric calibrations of each individual exposure are circled and labeled N1–N8 in green. The insert in the bottom left shows a single 360-seconds exposure of the lens. Reproduced from Bonvin et al. (2017).

4. The final step consists of deconvolving each exposure by the PSF previously modeled. Simply put, since the diffraction effects produced by the atmosphere and optical elements of the telescope are modeled by their respective PSF, then deconvolving the exposures by the PSF should in theory concentrate all the diffracted light into much smaller numbers of pixels, providing the PSF has been accurately modeled. In practice, since both the observations and the PSF models are noisy, the deconvolution will fatally be noisy as well and exhibit residuals. Of course, a proper deconvolution scheme should minimize these residuals, but there is a degeneracy between the possible deconvolutions and the associated minimizations, thus making the problem ill-posed. The solution adopted in the MCS deconvolution scheme is to add a *regularization* term, i.e. adding new extra information that reduce the space of plausible solutions. In practice, the regularization happens already at the stage of the PSF construction, by choosing a sampling for the PSF of twice the one of the original exposure, as explained in point 3. The deconvolution process is then executed simultaneously on all exposures; they share the same deconvolved model, which is decomposed into two channels: a collection of analytical point sources representing the stars and the lensed images of the quasar, and a numerical pixel channel representing the image of the lensing galaxy and any

other potential extended object in the vicinity. During the deconvolution, the position of the point sources and the structure in the numerical channel is iteratively adapted simultaneously for all exposures, but the intensities of the point sources is allowed to vary from exposure to exposure. In the end, the deconvolution yields a common *background* to all exposures, as well as point sources of common position but variable intensity.

Only cut-outs of each science exposure are deconvolved. First, a number of well-selected stars around the lens are chosen, supposedly of constant magnitude, in order to compute a relative normalization coefficient from exposure to exposure. In the case of HE0435-1223, these stars are labeled in green on Fig. 2.4. Second, the lensed images of the quasar are simultaneously deconvolved. The resulting normalized intensities of the deconvolved point sources are then used to draw the light curves of the lensed images. As an illustration of the deconvolution process, Fig. 2.5 presents various cut-outs produced by the pipeline that are used to visually check the products of the deconvolution.

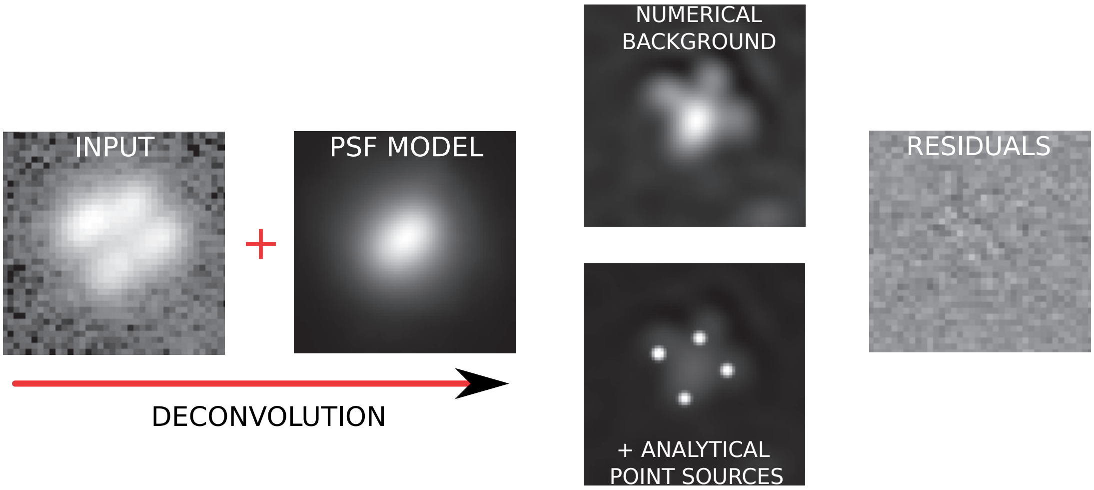


Figure 2.5: Cut-outs of the products of the COSMOULINE pipeline, made in order to visually assess the quality of the deconvolution process. The leftmost panel presents the "raw" input data prior to the deconvolution. The second-from-left panel presents the PSF model computed on the same exposure. Deconvolving the input data with the PSF model yields two different channels, represented in the third-from-left column: the numerical one (the background) representing the pixelated lens galaxy and extended source host galaxy structure, and the analytical one with the quasar images as point sources thousand times brighter than the reconstructed background. The right panel presents the residuals of the deconvolution.

Although the whole COSMOULINE pipeline is now largely automated, it still requires a non-negligible amount of human input when a new target is analyzed for the first time. Namely, the choice of PSF and normalization stars as well as the parameters of the MCS deconvolution scheme requires to be tested as extensively as possible. A schematic view of the main steps of the pipeline detailed above is presented in Fig. 2.6.

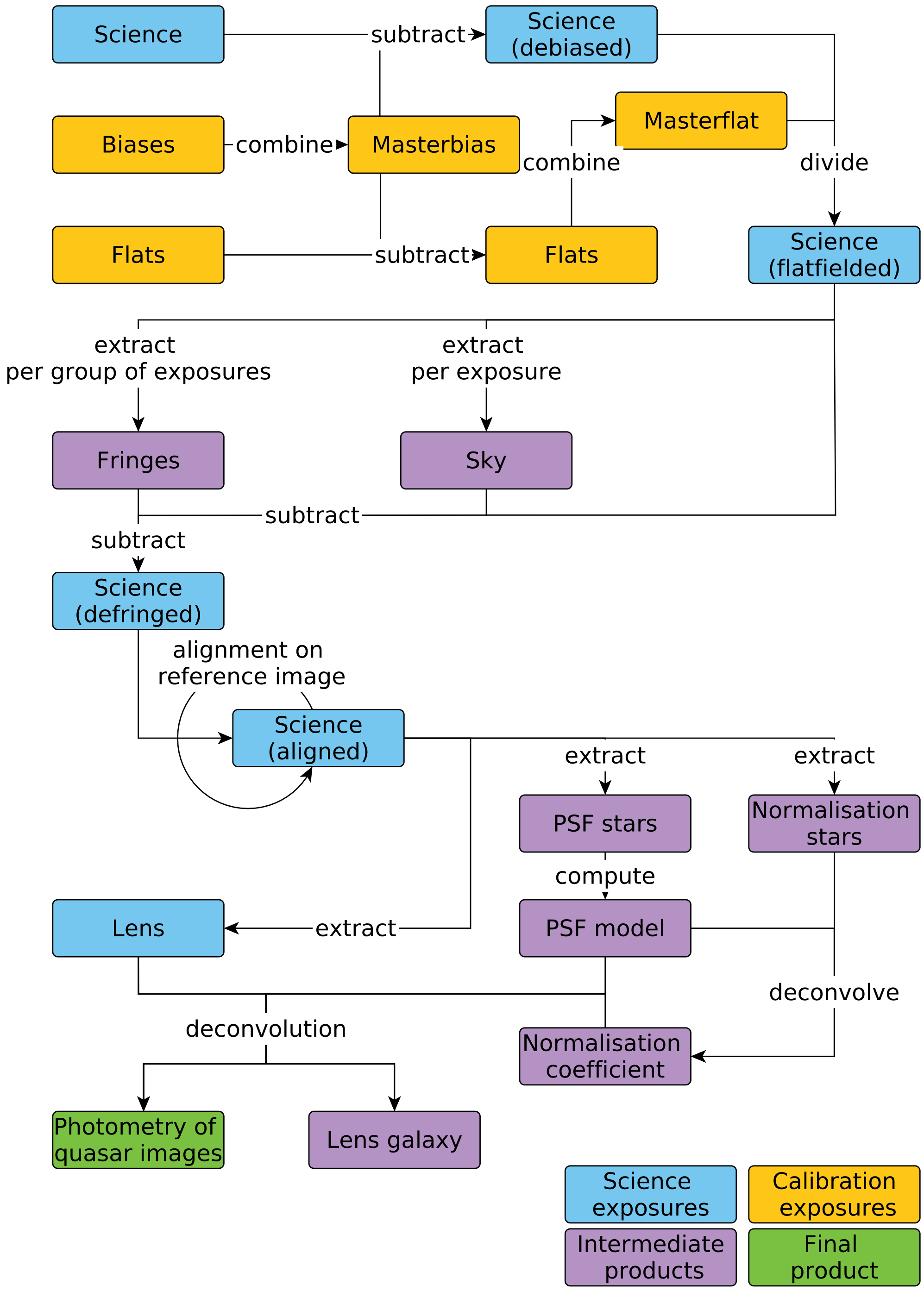


Figure 2.6: Schematic view of the COSMOULINE pipeline. The data reduction starts on top left, using the raw Science, Biases and Flats exposures.

2.4 Light Curves

All the data acquired on the targets presented in Table 2.1 have been at some point processed through the pipeline described above, which yielded for each target a set of light curves of the lensed quasar images. In this section, I present the light curves of a selection of the most interesting COSMOGRAIL targets.

2.4.1 SMARTS data

Before going further, it is important to mention that COSMOGRAIL faced, from 2000 to 2010, a concurrent collaboration. It is named the *Small and Moderate Aperture Research Telescope System* consortium (hereafter SMARTS[3]), and regroups many American institutes. SMARTS makes use of four 1-meter class telescopes located at the Cerro Tololo Interamerican Observatory (CTIO), in Chile. The research performed by SMARTS covers a wide variety of topics, including the monitoring of lensed quasars (see e.g. Kochanek et al., 2006; Fohlmeister et al., 2008). However, since the interest of the people involved in strong gravitational lensing in SMARTS is more on the characterization of the microlensing events than on the measurement of time delays, it has been agreed that SMARTS and COSMOGRAIL should collaborate and share their data on the targets they have in common. In practice, it means that I had access to some of the already flat-fielded data from the 1.3-meter optical ANDICAM instrument. By processing them through COSMOULINE, I was able to produce light curves of a quality comparable to the other COSMOGRAIL telescopes. This allowed me to seamlessly merge the COSMOGRAIL and SMARTS data into a single set of light curves, enhancing the baseline and the sampling. In what follows, I present the merged data as COSMOGRAIL light curves, without distinction between the instruments used.

2.4.2 HE0435-1223

HE0435-1223 has been the main object I analyzed during the course of my PhD thesis. It is a bright quad in a symmetric configuration, visible in Fig. 2.2 and Fig. 2.4. HE0435-1223 (α(2000): 04h38m4.9s; δ(2000): -12°17'14"4) has been discovered by the Hamburg/ESO Survey (HES) for bright quasars in the Southern hemisphere (Wisotzki et al., 2000, 2002). The redshift of the lens has been independently reported by Morgan et al. (2005) and Eigenbrod et al. (2006b) as $z_{\rm d} = 0.4546 \pm 0.0002$ and the most recent determination of the source redshift is reported by Sluse et al. (2012b) as $z_{\rm s} = 1.693$. The lens lies in a small cluster that contains at least 12 galaxies (Sluse et al., 2016). The four lensed images have a relatively similar absolute luminosity, with a difference smaller than 1 magnitude between the brightest (A) and the faintest (D) image. The COSMOGRAIL light curves have been first presented in (Courbin et al., 2011), and then updated in (Bonvin et al., 2017) with six more years of data from ECAM. They are reproduced in Fig. 2.7.

---

[3] http://www.astro.yale.edu/smarts/

2.4.3 RXJ1131-1231

RXJ1131-1231 has been the main object analyzed during the previous COSMOGRAIL thesis (Tewes et al., 2013c). It is another bright quasar, in a cusp configuration, (α(2000): 11h31m52s; δ(2000): -12°31’59"), serendipitously discovered during imaging observations of radio quasars in 2002 (Sluse et al., 2003). The redshift of the source and the lens have been measured by Sluse et al. (2003) as $z_s = 0.658$[4], and $z_d = 0.295$, respectively. The difference in magnitude between the brightest (A) and faintest (D) images is up to 4 magnitudes. The COSMOGRAIL light curves have been first presented in (Tewes et al., 2013b). Figure 2.8 presents the updated version with four more years of data.

---

[4] Later revised as $z_s = 0.654$, as used in Bonvin et al. (2017)

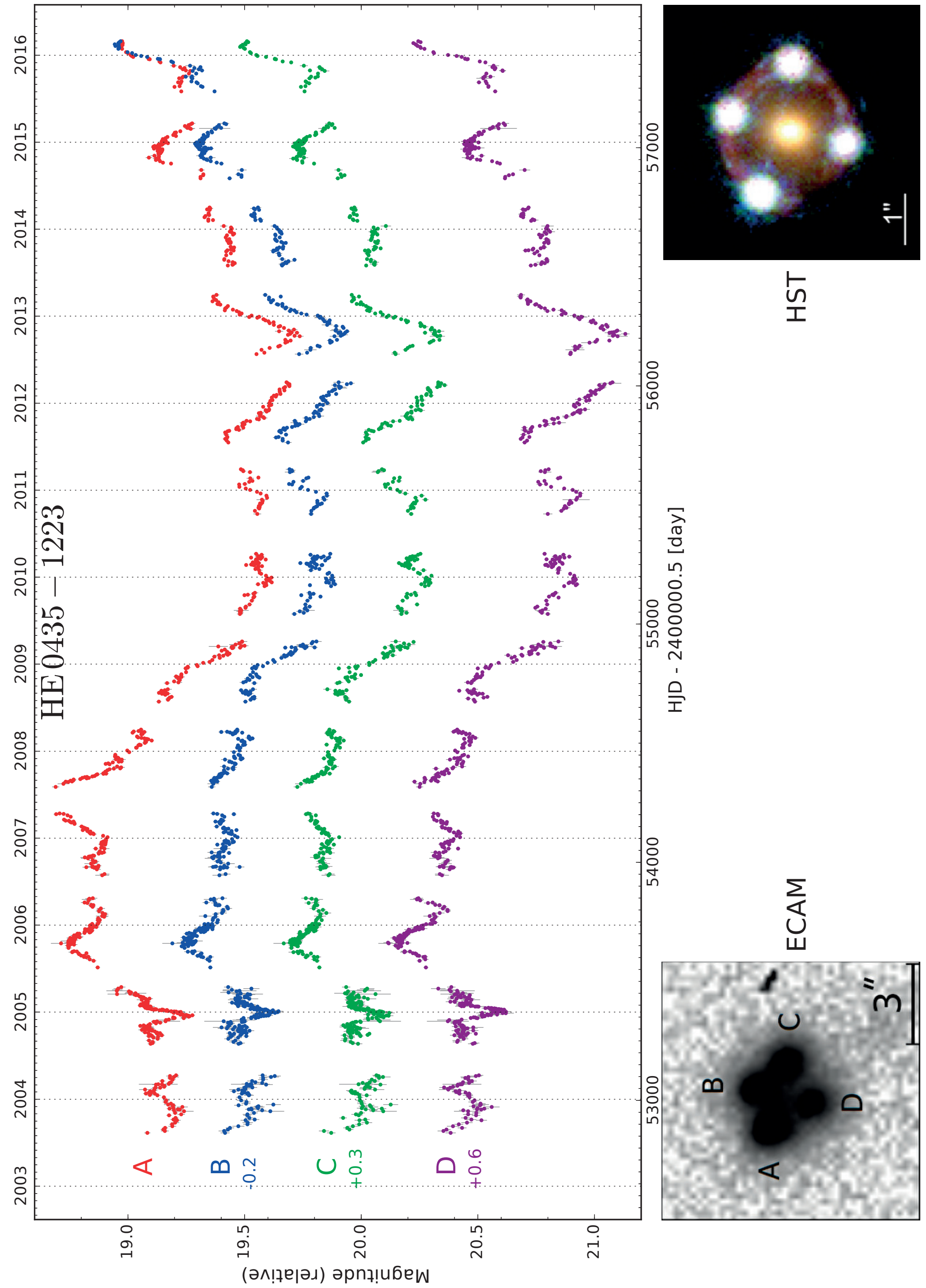


Figure 2.7: 13-years light curves of the lensed quasar HE0435-1223, adapted from Bonvin et al. (2017). The bottom panels show a single 360-second ECAM exposure of the quasar (left) and an combined color HST exposure adapted from Suyu et al. (2016) (right). Note that the lens galaxy and Einstein ring appear in the HST image but are absent from the ECAM exposure.

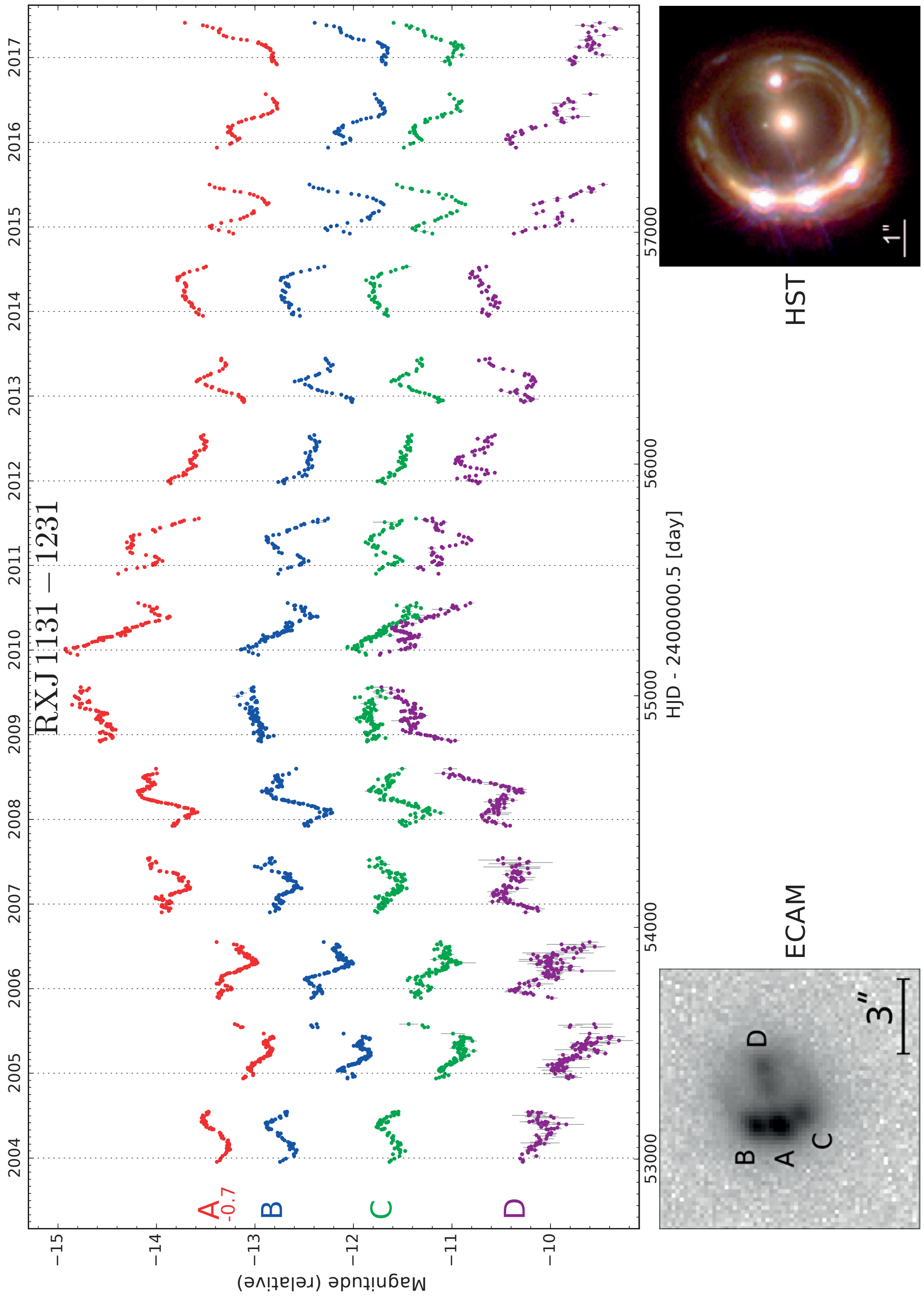


Figure 2.8: 13-years light curves of the lensed quasar RXJ1131-1231. The bottom panels show a single 360-second ECAM exposure of the quasar (left) and an combined color HST exposure adapted from Suyu et al. (2016) (right). The Einstein ring, barely visible in the ECAM cut-out is seen very distinctly in the HST one. The blue regions seen inside the Einstein ring are different parts of the host galaxy magnified by the lens (Paraficz et al., 2017).

2.4.4 WFI2033-4723

WFI2033-4723 is one of the two targets of the H0LiCOW sample (see Sec. 4.2) that is currently being thoroughly analyzed for time-delay cosmography (Suyu et al., 2016). This lensed quasar has been discovered by Morgan et al. (2004) during an lens survey in the optical conducted at the ESO/MPG 2.2m telescope in La Silla Observatory. It is a quad in a fold configuration (α(2000): 20h33m42.08s; δ(2000): -47°23'43"). The most recent determination of the source and lens redshifts yield $z_s = 1.662$ (Sluse et al., 2012c) and $z_d = 0.661$ (Eigenbrod et al., 2006b). The lens apparently lies in a group of galaxies, whose redshift still need to be properly estimated (Suyu et al., 2016). The COSMOGRAIL light curves have been first presented in (Vuissoz et al., 2008). The four lensed images have the remarkable property to have nearly the same apparent magnitude. Yet, A1 and A2 being really close to each other (less than 1" apart), it is extremely difficult to resolve them properly in ground-based observations, even when using a proper deconvolution scheme like MCS. Luckily, the lens configuration allows us to make the reasonable assumption that the time delay between the two merging images is very small, likely compatible with zero. Thus, to avoid any flux sharing issues when drawing the light curves, we chose to sum the flux of A1 and A2 into one single curve, simply called A. Figure 2.9 presents the updated version of the COSMOGRAIL light curves with ten additional years of data, i.e. more than quadrupling the duration of the monitoring campaign of Vuissoz et al. (2008).

2.4.5 HE1104-1805

HE1104-1805 is the second of the two targets of the H0LiCOW sample that is currently scrutinized by two dozen astronomers (Suyu et al., 2016), and the only double in the lot. It has been discovered by the Hamburg/ESO Survey by (Wisotzki et al., 1993) (α(2000): 11h06m33.39s; δ(2000): -18°21'23.8"). The source and lens redshifts have been determined respectively as $z_s = 2.316$ (Smette et al., 1995) and $z_d = 0.729$ (Lidman et al., 2000). Five years after the initial discovery of the lensing images, Courbin et al. (1998) and Remy et al. (1998) both independently observed the lens galaxy for the first time in the infra-red band, surprisingly showing that the brighter lensed image lies closer to the lens galaxy - a possible yet unusual configuration. Figure 2 of Suyu et al. (2016) shows that even after a crude subtraction of the lensed images, arc-like structures become visible around the lens galaxy. The large separation between the two lensed images (around 3") as well as their apparent brightness makes HE1104-1805 a very promising system for time-delay cosmography. No COSMOGRAIL light curves have been published prior to this thesis, although the first two years of SMARTS monitoring were presented in Poindexter et al. (2007). Figure 2.10 presents 13 years of monitoring using mainly SMARTS data (2004 to 2016) completed with three years of ECAM data (2013 to 2016).

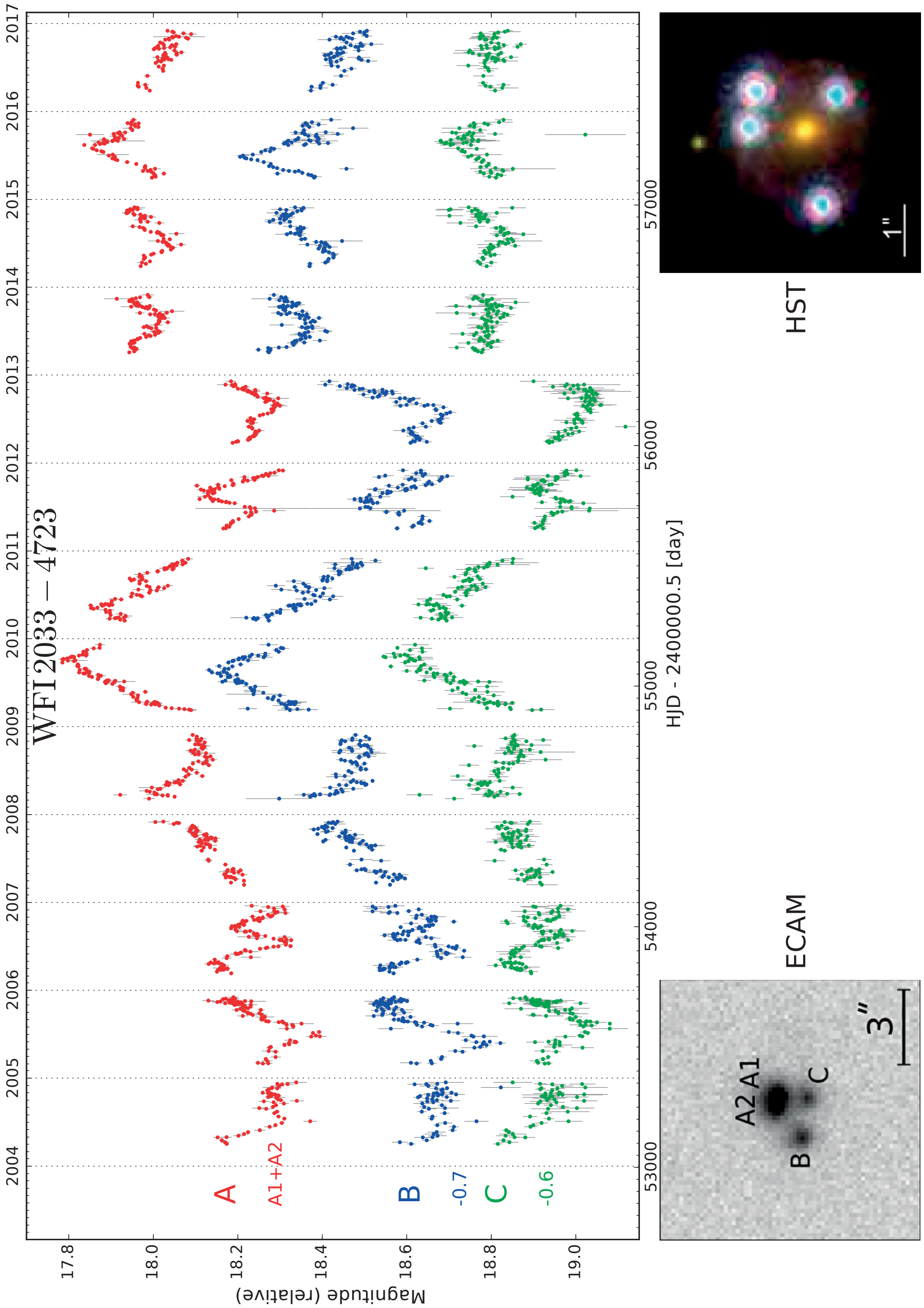


Figure 2.9: 13-years light curves of the lensed quasar WFI2033-4723. The bottom panels show a single 360-second ECAM exposure of the quasar (left) and a combined color HST exposure adapted from Suyu et al. (2016) (right).

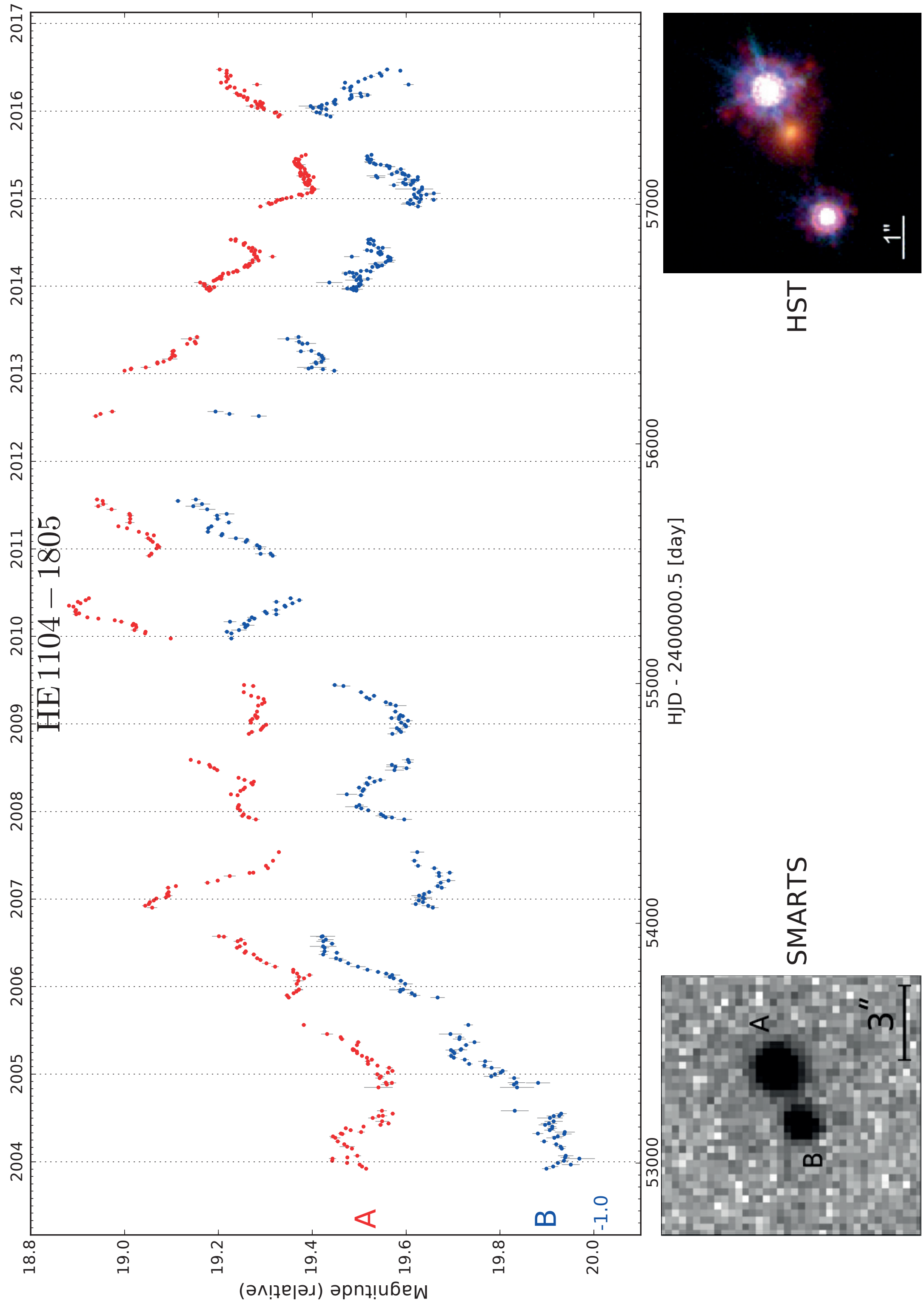


Figure 2.10: 13-years light curves of the lensed quasar HE1104-1805. The bottom panels show a single 300-second SMARTS exposure of the quasar (left) and a combined color HST exposure adapted from Suyu et al. (2016) (right).

2.4.6 HE0047-1756

HE0047-1756 has been discovered by the Hamburg/ESO Survey by (Wisotzki et al., 2004) (α(2000): 00h50m27.83s; δ(2000): -17°40'08.8") through multi-band imaging at the 6.5m Magellan telescope, at a redshift of $z_s = 1.67$. The redshift of the lens was measured at the VLT by Eigenbrod et al. (2006b) and Ofek et al. (2006), leading to $z_l = 0.407$. HE0047-1756 has been observed for 13 years at Euler. It is also one of the first target monitored at high cadence with the 2m2, providing some interesting results despite the very short amount of time dedicated to it (only three months, see Sec. 2.5.3). I reproduce in Fig. 2.11 the 13-years light curves from the Euler data. Note that recently, Giannini et al. (2017) published light curves from 5 years of optical data taken at the 1.5m Danish telescope located at La Silla.

2.4.7 Other lenses

In addition to the five lenses presented above that have recently been and/or will be in a near future presented in publications, I also present here the light curves of some additional targets of the COSMOGRAIL database. The data reduction has been conducted in collaboration with Eric Paic, a summer student at the Laboratory of Astrophysics of EPFL. These reductions were carried out in order to have a first quick look at the light curves, mainly to assess if it was worth keeping on with the monitoring. In such a process, various steps of the data reduction have not been optimized. Namely, the choice of the best PSF and normalization stars, as well as the modeling parameters of the MCS deconvolution have not been fully explored. Once first apparently satisfying setups were found, they were used without further investigation. Obviously, in the scope of a publication in a scientific paper, such an approach would not be sufficient. That must be kept in mind when looking at the light curves that follow.

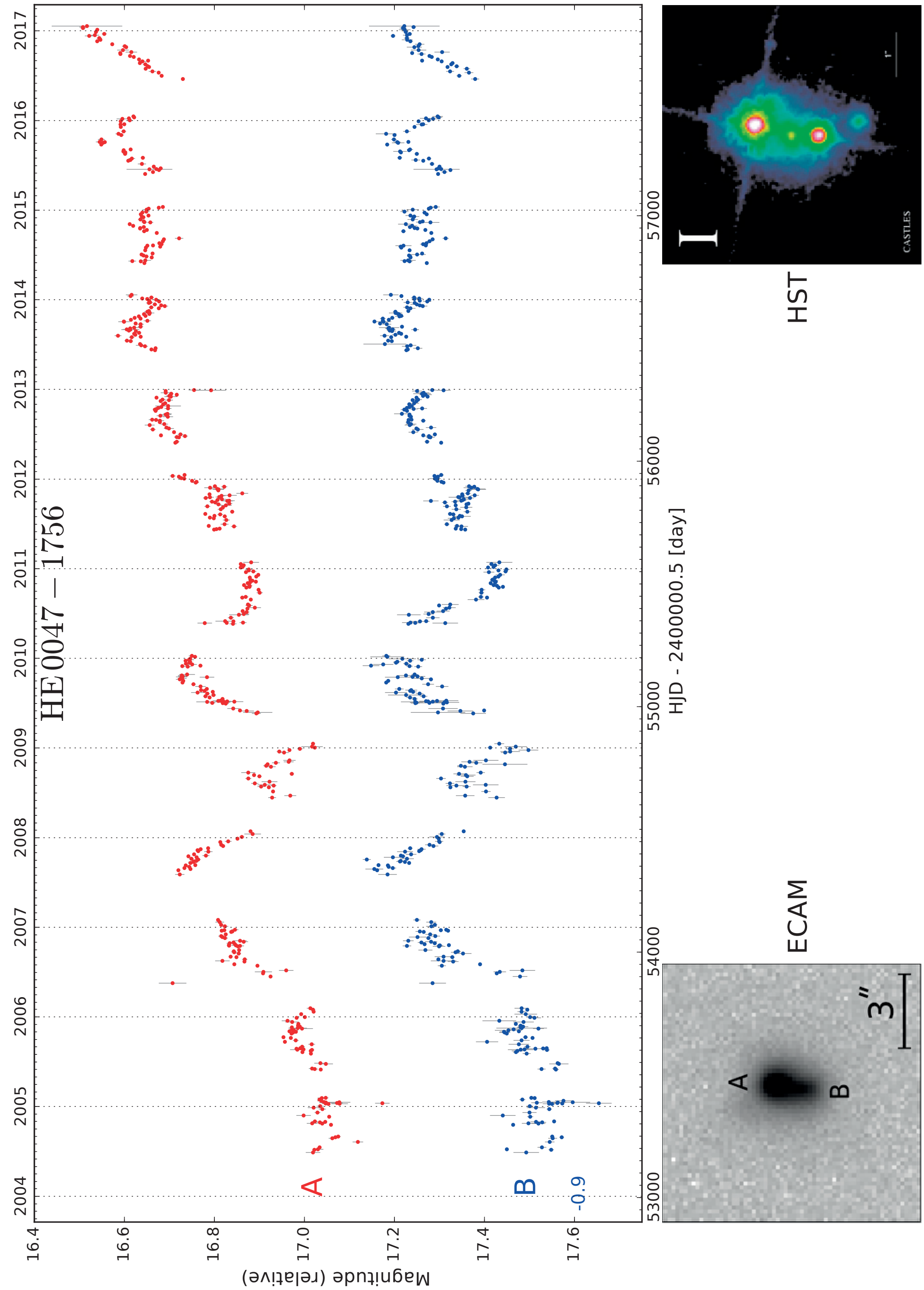


Figure 2.11: 13-years Euler light curves of the lensed quasar HE0047-1756. The bottom panels show a single 360-second ECAM exposure of the quasar (left) and an I-band HST exposure adapted from Kochanek et al. (1998) (right).)

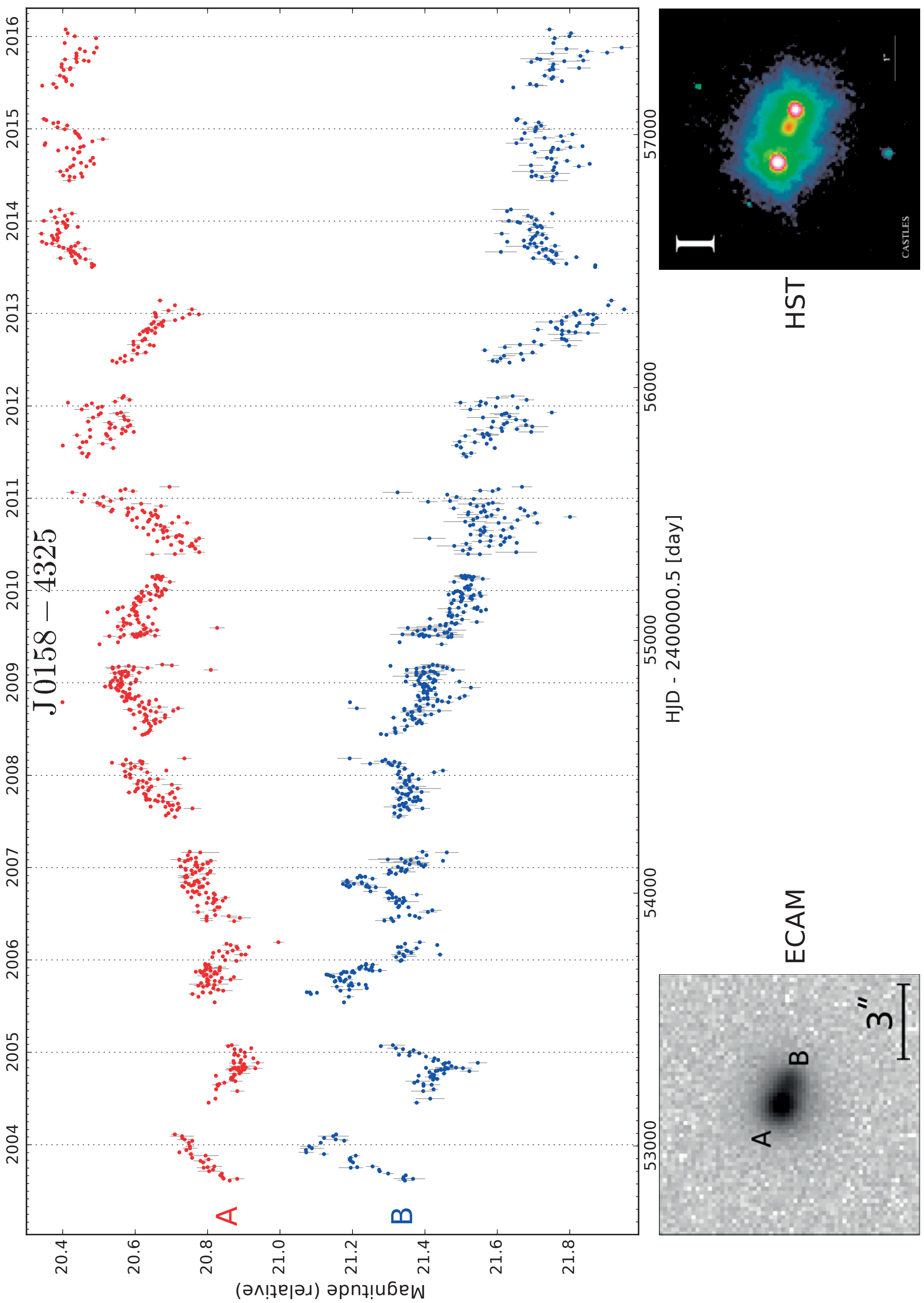


Figure 2.12: 13-years Euler light curves of the lensed quasar J0158-4325. The bottom panels show a single 360-second ECAM exposure of the quasar (left) and an I-band HST exposure adapted from Kochanek et al. (1998) (right).

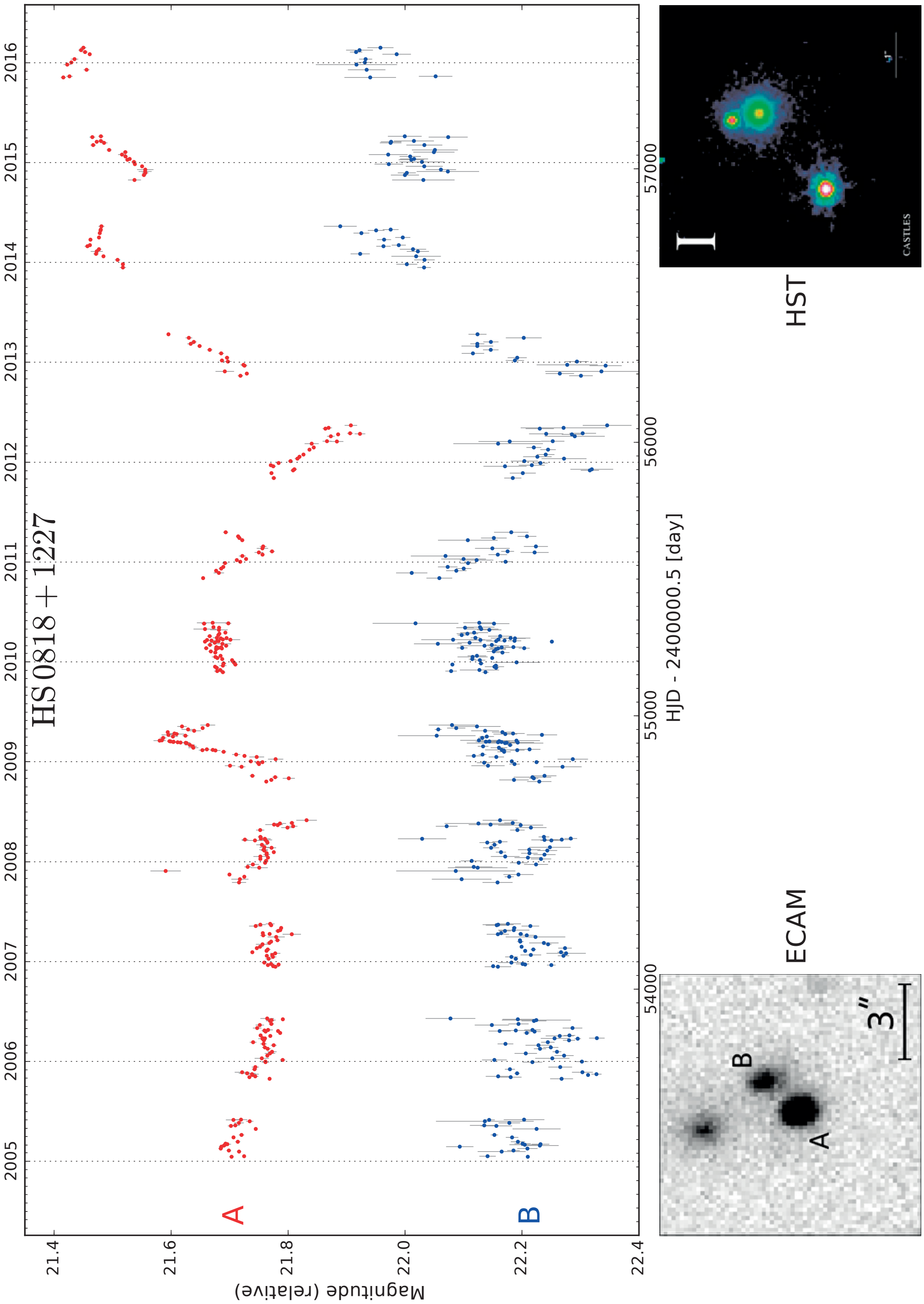


Figure 2.13: 12-years Euler light curves of the lensed quasar HS0818+1227. The bottom panels show a single 360-second ECAM exposure of the quasar (left) and an I-band HST exposure adapted from Kochanek et al. (1998) (right).

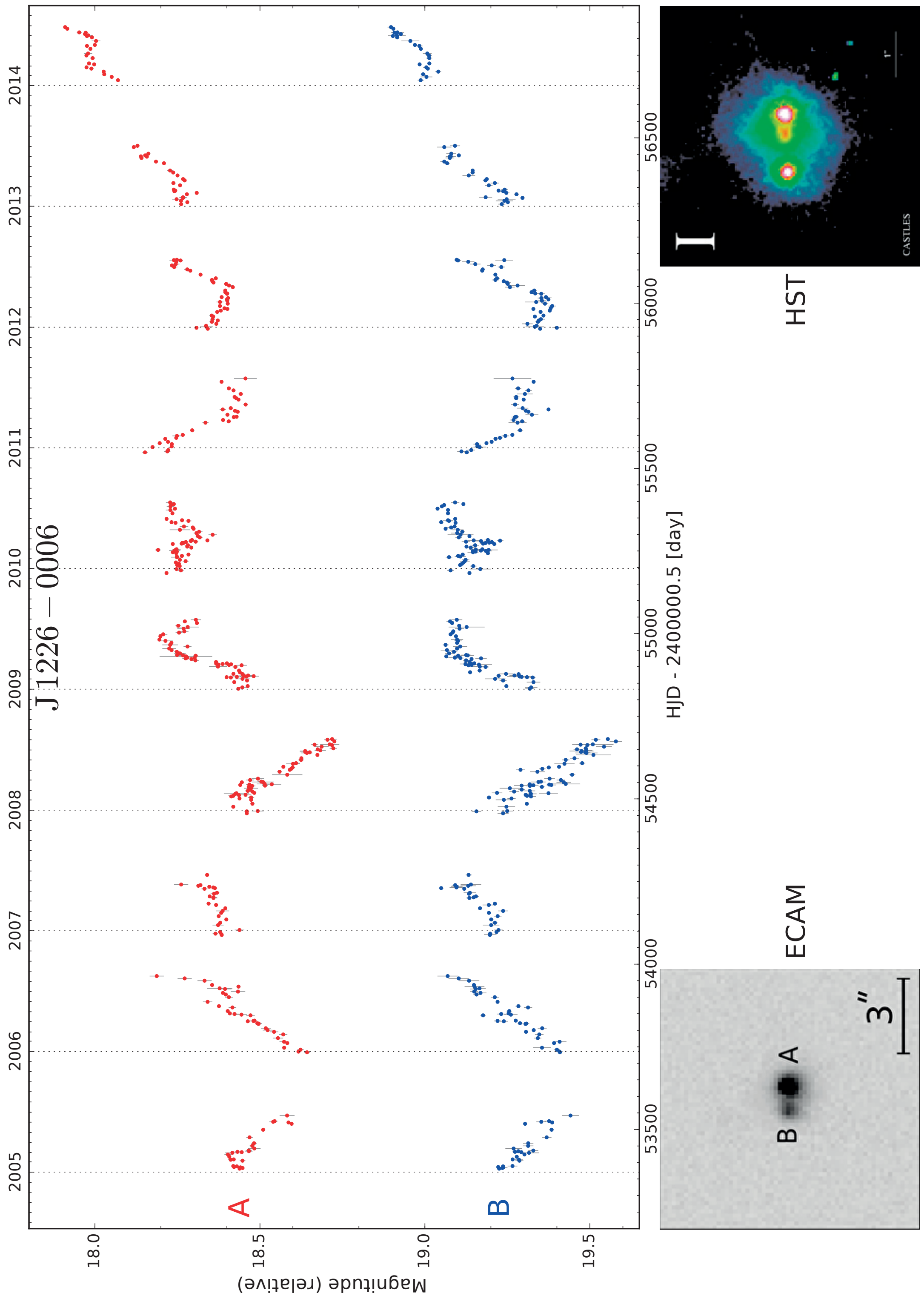


Figure 2.14: 10-years Euler light curves of the lensed quasar J1226-006. The bottom panels show a single 360-second ECAM exposure of the quasar (left) and an I-band HST exposure adapted from Kochanek et al. (1998) (right).

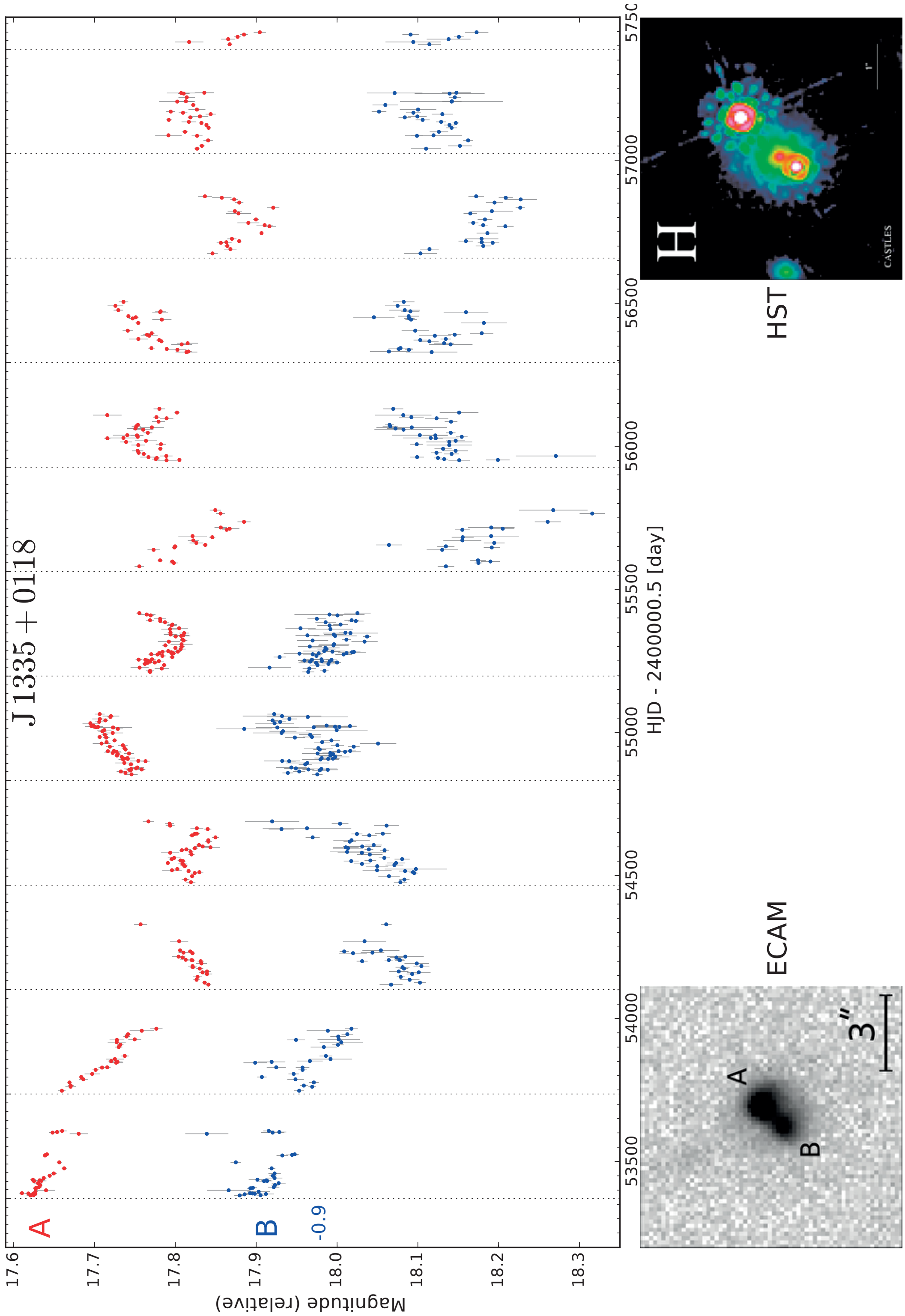


Figure 2.15: 11.5-years Euler light curves of the lensed quasar J1335+0118. The bottom panels show a single 360-second ECAM exposure of the quasar (left) and an H-band HST exposure adapted from Kochanek et al. (1998) (right).

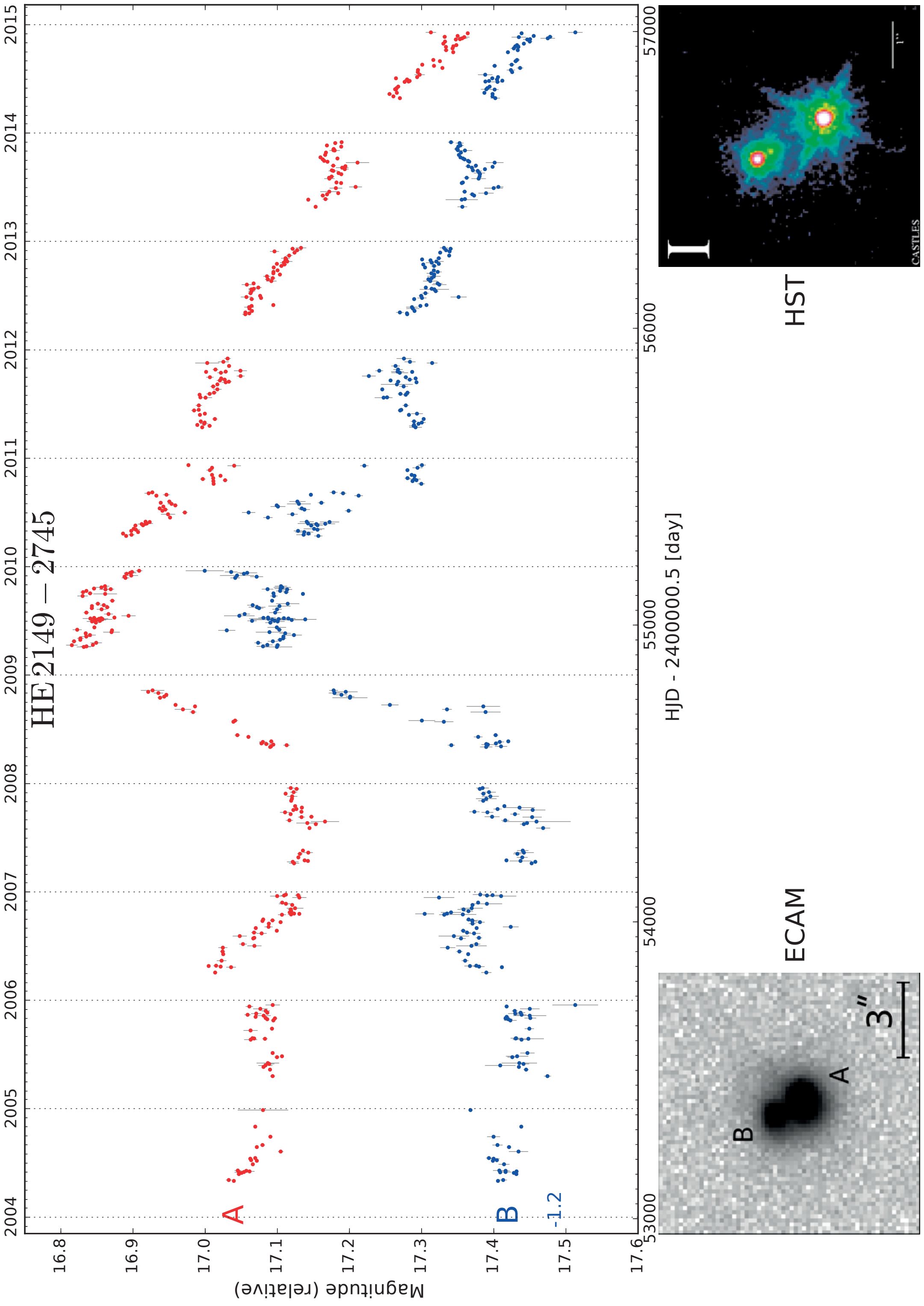


Figure 2.16: 11-years Euler light curves of the lensed quasar HE2149-2745. The bottom panels show a single 360-second ECAM exposure of the quasar (left) and an I-band HST exposure adapted from Kochanek et al. (1998) (right).

2.5 Speeding-up the measurements of time delays

In a few words, the idea is to monitor the lensed quasars with a bigger telescope, and at a higher cadence. Although it might sounds like a naive increment on an already existing program, it represents in reality a huge step forward in the field of time-delay measurements. Its scope can be understood through two main observations on the current state of research:

1. The COSMOGRAIL collaboration has successfully demonstrated in the past decade its ability to measure time delays with high precision and accuracy. Following the sometimes unprecise and conflicting time-delay estimates in the 1980's and 1990's, COSMOGRAIL established itself as the leader in the field, producing light curves and time delays that could be trusted and used to do precision cosmology.
2. The previous years saw the emergence of large sky surveys that started to discover more and more lensed quasars. For example, the Dark Energy Survey collaboration (DES, Dark Energy Survey Collaboration et al., 2016) has already discovered candidates that have been integrated to the COSMOGRAIL sample (e.g. DES0408-5354, Lin et al., 2017; Agnello et al., 2017). In combination with other ongoing surveys like the Kilo-Degree Survey (KiDS, de Jong et al., 2013), the VST-ATLAS Survey (Shanks et al., 2015) or the Hyper Suprime-Cam Survey (HSC, Chan et al., 2016), hundreds of new lensed quasars are expected to be discovered by the end of the decade. And looking a little bit more ahead, the upcoming Large Synoptic Survey Telescope (LSST, LSST Science Collaboration et al., 2009) or the Euclid mission (Laureijs et al., 2011) are expected to increase that number by at least one order of magnitude.

Obviously, not every lensed quasar that is going to be discovered will be suitable for time-delay cosmography. Yet, even a tiny fraction of the potentially 100'000 new strong lenses to be discovered would represent a significant increase. In order to make use of that great wealth of data, waiting up to a decade to get time-delay measurements precise and accurate enough is not a viable option anymore. A precise determination of the time delay being what usually takes the longest in time-delay cosmography, doing it in only one or two years after the discovery of the quasar would represent a significant step forward. We aim at addressing this step by adopting a new monitoring strategy, in order to get robust time-delay estimates in only one monitoring season instead of a decade.

Doing so requires to catch many small-scale variations in the luminosity of the lensed quasar images. Such variations are not only shorter but also much fainter, around one order of magnitude smaller than what COSMOGRAIL is currently sensitive to. Achieving this goal requires both to increase the photometric precision of the measurements and the monitoring cadence. With the current COSMOGRAIL small-size telescopes, reaching a sufficient photometric precision (in the order of a mili-magnitude) would require hours of exposition per target, which *de facto* drastically limits the number of observable targets per instrument. The obvious solution is thus to use bigger instruments. Nevertheless, it is not that easy to achieve since obtaining access to larger facilities either face strong competition or is terribly expensive. Fortunately, the good reputation that COSMOGRAIL acquired over the years helped a lot in that regard, easing the call for human and financial resources to properly set up the project.

2.5.1 A pilot monitoring campaign of HE0435-1223 using the Very Large Telescope

To assess the feasibility and performances of the project, a pilot monitoring campaign was conducted with one of the Very Large Telescopes operated by ESO in Paranal, Chile. The instrument used, called FORS2 (for FOcal Reducer/low dispersion Spectrograph 2), provides a 6.8 x 6.8 arcmin field-of-view in imaging mode, comparable with Euler in that regard. However, it is mounted on a 8-meter diameter telescope, which collects much more luminosity in a shorter amount of time and thus allow to reach the required photometric precision in a few minutes instead of hours of exposure. The planned monitoring cadence was of one observation per night, on the lensed quasar HE0435-1223, continuously during six months, with a few undesired yet impossible to avoid chunks of nights without observations due to maintenance of the instrument. HE0435-1223 also continued to be monitored at Euler during these six months, with an increase of the monitoring frequency from one observation every four night to one observation per night during the maintenance periods of FORS2. The left panel of Fig. 2.17 presents the resulting light curves of both FORS2 and ECAM during the whole campaign, from October 2014 to March 2015. To illustrate the difference of sensitivity between the two instruments, the right panels display two single exposures of HE0435-1223, taken in similar observing conditions.

A striking characteristic of the FORS2 light curves, that would have been happily avoided if possible, is the highly uneven sampling. This results from two main factors. First, the automatic observation scheduler used at the VLT to help the astronomers on-site to prepare their observing night was not designed to properly handle monitoring programs. In such cases, especially during the first months, scheduling conflicts sometimes prevented the lensed quasar observations to be routinely performed, although the observing conditions were optimal. In addition, technical nights and visitor mode[5] were more numerous than expected on our side, leading to more nights and/or chunks of nights without observations. Second, the weather has been sensitively worst than expected. Paranal has originally been chosen as a site to build the VLT because of its excellent observing conditions, with only a few nights per year without good enough conditions to observe. Yet, starting roughly from July 2014 to October 2016 the weather conditions in Chile worsened and impacted most of the observing programs, including ours[6]. Obviously, our monitoring campaign that was particularly sensitive to the sampling suffered a lot from these combined effects, that are directly visible in the light curves.

---

[5] The VLT goes into *visitor mode* when an astronomer travels up to Paranal in order to make sure his/her observations are properly carried out by the technical staff. In such cases, we learned to our great surprise that the visitor's observations got the priority over the monitoring campaign.

[6] For the record, spring 2016 has been designed as the worst observing period since the beginning of ESO, with less than 30% of the planned observations carried out.

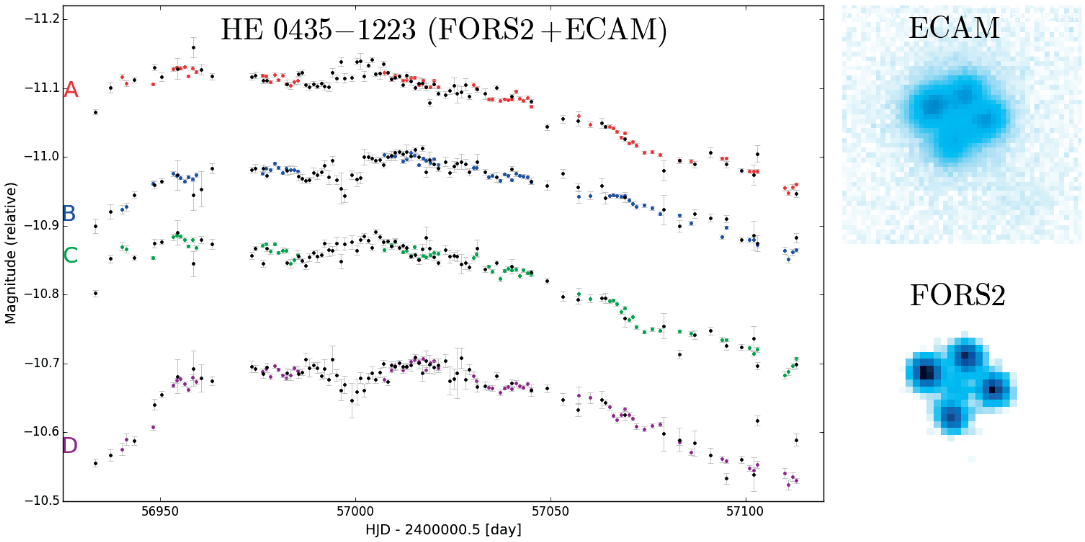


Figure 2.17: The left panel presents the light curves of HE0435-1223 monitored with FORS2 (colored points) and ECAM (black points). The data span a total of six months, from October 2014 to March 2015 both months included. The two right panels present cut-outs of single exposure of the lensed quasar, seen with the two instruments: on top, a 360-seconds ECAM exposure and at the bottom, a 12-seconds VLT exposure. Both exposures were taken in similar observing conditions, with a DIMM seeing of ~0.8".

Two other significant things to note in the FORS2 light curves are the strong night-to-night scatter of the measurements with respect to the expected photometric mili-magnitude precision, as well as the correlated noise in the light curves (i.e. features appearing simultaneously in the four light curves, not delayed in time). This is harder to detect when looking at Fig. 2.17, but is nonetheless an issue that we suffered from. To understand where it comes from, let us go back to COSMOULINE pipeline (Sec. 2.3.1). The photometric errors associated to the measurements as computed in the pipeline are a combination of the Poissonian shotnoise and empirical intra-night normalization coefficient computed over all the exposures taken during the same night. The latter is computed using various reference stars in the field, whose luminosity is supposed to be stable over time. Thus, if for some reason the luminosity of the reference stars vary from exposure to exposure, the photometric error on the measured magnitude of the lensed images is adapted accordingly. However, no cross-checks are performed between different observing nights. As a result, if the normalization stars vary on a period of a day or more due to either physical or instrumental causes, the resulting light curves would be affected, but not the individual night photometric precision.

A known issue of FORS2 is that the Longitudinal Atmospheric Dispersion Corrector (LADC), due to a degradation of its coating, produced a spurious low-frequency pattern in the flat field that depends on the rotation angle between the LADC and the telescope filters. And that angle, although stable during a series of consecutive exposure, could completely change from night to night. That effect has been deeply studied (Moehler et al., 2010; Coccato et al., 2014), and a proposed fix to it has been integrated into the calibration plan and standard data reduction procedure of FORS2. In late 2014, a change of the LADC to another one with a much less severe coating degradation sensibly improved the overall photometric precision of FORS2 to 1% (Boffin et al., 2015, 2016), sensitively better than what could be reached through the previously proposed fix. As a result, the latter has been removed from the standard calibration plan of FORS2. Yet, although the papers above claim that these changes allow to use FORS2 for precision photometry again, it was definitely not sufficient for the mili-magnitude precision we needed to achieve.

Taken separately, the issues mentioned above (non-optimal schedule, bad weather, LADC, ...) could certainly have been handled. Yet, taken altogether they affect the resulting light curves to a point that is just not good enough to precisely measure a time delay (see Sec. 3.4.1). Bluntly said, that pilot monitoring campaign with the VLT was a failure. Yet, it provided plenty of useful insights that we used to better design the next phases of the high-cadence monitoring mode.

2.5.2 Monitoring with the 2.2m ESO/MPG Telescope

The pilot VLT monitoring campaign not having been a success, it was important to learn what should be improved before moving on. If it is anyway impossible to avoid bad weather without using a space telescope, far too expensive, the other issues were potentially all avoidable. All considered, it reduces to two main criteria. First, the optical assembly of the telescope should be as simple as possible and the electronics as clean as possible to ensure that when reaching for high photometric precision, the limiting factor stays the photometric shot noise of the observations. Second, the nightly observation window of the monitoring program must be wide enough so that even if half of the night is lost for technical reasons, other high-priority observations or bad weather, the planned targets could still be observed. Indeed, simulations performed after the VLT campaign (Courbin and Bonvin, prep, detailed later on in Sec. 2.5.6) show that a high and regular sampling cadence is one of the most important criteria for a successful monitoring campaign. Telescope and budget-wise, we had two options: the 3.6m New Technology Telescope (hereafter NTT) and the 2.2m ESO/MPG Telescope (hereafter 2m2), both located at La Silla, in Chile. After a few nights of test observations, the latter was finally chosen. Despite being smaller, its simpler optics and higher availability made it more suitable to our needs.

We started the observations on the 1st of October 2016, making use of of the Wide-Field Imager instrument (hereafter WFI) mounted on the 2m2. WFI has a total field of view of 36x36 arcminutes, covered by 8 CCDs with a pixel size of 0.24 arcseconds. We make use of only one of these chips to ensure a stable night-to-night calibration. As an illustration, figure 2.18 presents a zoom on the field-of-view of the lensed quasar DES J0408-5354 seen with WFI.

Currently, 1.5 hours are dedicated to the monitoring during Chilean Summer time (October to March) and 2.0 hours during Chilean Winter time. Divided by the necessary time to reach the

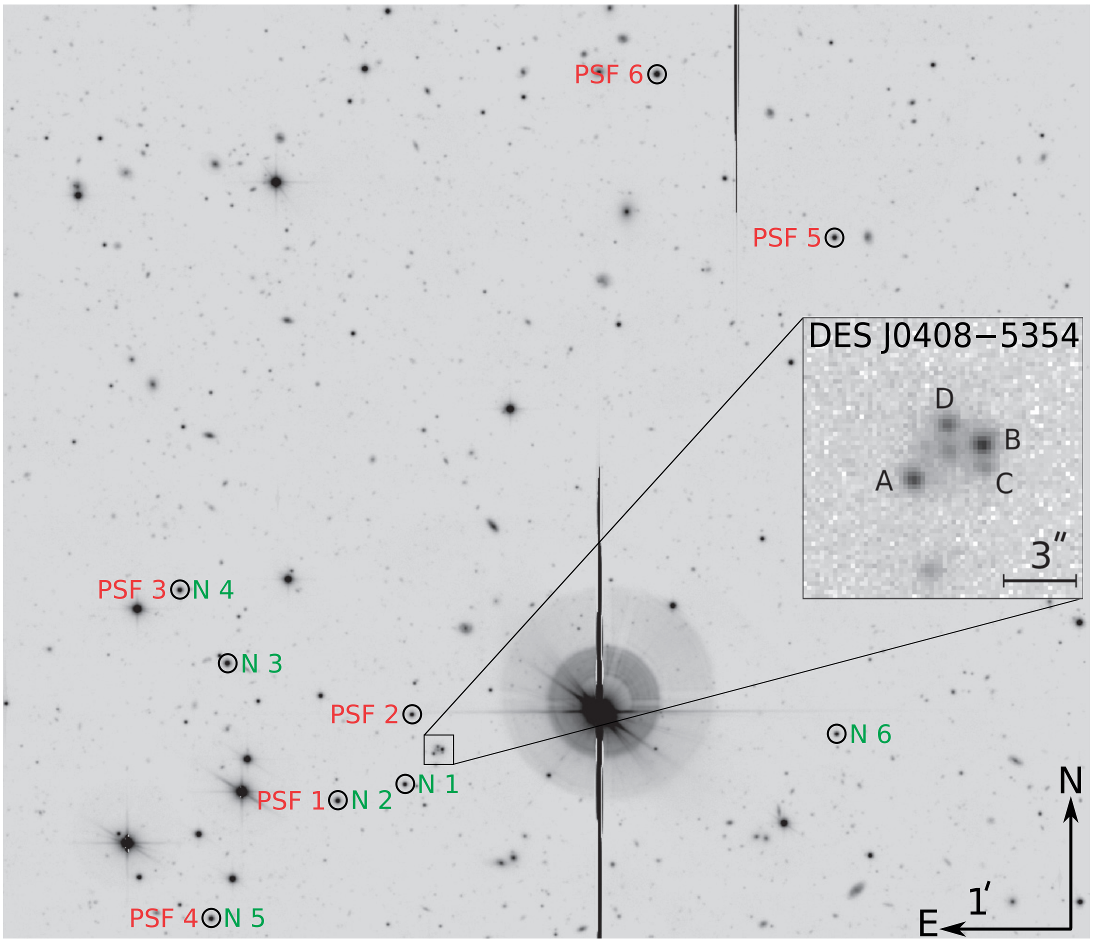


Figure 2.18: Part of the field-of-view of DES J0408-5354 seen through the WFI instrument mounted on the 2m2 telescope - roughly 1/20th of the whole field. This is a combination of 92 exposures of ~300 seconds each, for a total exposure time of ~7 hours. The stars used to build a PSF model are circled and labeled PSF 1 to PSF 6 in red, and the stars used for the photometric calibrations of each individual exposure are circled and labeled N 1–N 6 in green, as detailed in Sec. 2.3.1. The insert on the middle right shows a single 300-second exposure of the lens. Reproduced from Courbin et al. (2017).

mili-magnitude photometric precision with a 2.2-metre mirror, we scheduled the monitoring of three lenses per night in Winter time, and four in Summer time. The calendar of observations for the first year of monitoring at the 2m2 is presented in Tab. 2.2.

The choice of targets for the first monitoring year at the 2m2 is balanced between "old" and "new" lenses. On one hand, monitoring bright quasars already well studied in the past decades like PG1115+080 is a good test to assess the achievability and precision of the program. On the other hand, freshly discovered targets like DES J0408-5354 and DES2038-4008 are perfect cases to demonstrate the efficiency of the program, the goal being to come up with a precise and accurate time delays in no more than two seasons of monitoring. Finally, targets like

Table 2.2: List of all the targets being monitored at the 2.2m ESO/MPG Telescope during the 1st year of the monitoring campaign (from 1st October 2016 to 1st October 2017), along with their discovery date, most recent published time-delay estimate for the longest delay and their monitoring period. A new target will replace HE1104-1805 starting August 2017, but its identity has not been decided at the time this document is written.

Name	Discovery	Measured time-delay [day]	Monitoring period
HE0047-1756	Wisotzki et al. (2004)	$\Delta t_{AB} = 7.6 \pm 1.8$ (Giannini et al., 2017)	01/10/2016 15/01/2017 —— 01/06/2017 01/10/2017
DES J0408-5354	Lin et al. (2017)	–	01/10/2016 31/03/2017
DES2038-4008	2017 (DES, private comm.)	–	01/04/2017 01/10/2017
PG1115+080	Weymann et al. (1980)	$\Delta t_{BC} = 17.6 \pm 6.9$ Tsvetkova et al. (2016)	15/12/2016 11/07/2017
HE1104-1805	Wisotzki et al. (1993)	$\Delta t_{AB} = 152^{+2.8}_{-3.0}$ Poindexter et al. (2007)	15/11/2016 30/07/2017
WFI2033-4723	Morgan et al. (2004)	$\Delta t_{BC} = 62.6^{+4.1}_{-2.3}$ Vuissoz et al. (2008)	15/03/2017 01/10/2017

HE0047-1756, HE1104-1805 and WFI2033-4723 are in between. Their potential for time-delay cosmography is high, with already good acquired follow-up data - the latter two targets being also in the H0LiCOW sample. With already known time delays and a long baseline of existing COSMOGRAIL and SMARTS data, the 2m2 measurements will be at least complementary and at best more precise, all this in less than a year.

The expectations were that after a few months of monitoring, one must be able to clearly see variability in the light curves of the various objects and measure time delays. At the time this manuscript is written, the year-one planned observations were completed for four targets: HE0047-1756, DES J0408-5354, PG1115+060 and HE1104-1805. They lasted respectively three, six, seven and eight months, spanning the visibility windows of the four targets during which they were high enough above the horizon to be properly observed. The so acquired exposures of the three former targets were processed through the COSMOULINE pipeline. Little adjustments relative to the 2m2 telescope were required, notably to include the subtraction of the fringes presented in Fig. 2.3. The resulting light curves are presented below.

2.5.3 HE0047-1756

HE0047-1756, already monitored for more than a decade with ECAM, was originally planned to be observed continuously during three months and a half with the 2m2. However, a technical issue with the camera prevented the observations to be carried out during ~20 nights, from the 12th of December 2016 until the end of that year. The monitoring started again in early 2017 but lasted only 15 nights until the target became not observable anymore. Since the expected time delay was roughly of the same duration, no observed structure in the leading light curve could be related to their equivalent in the trailing curve during these extra two weeks in 2017. The monitoring was carried on anyway, just in case the measured time delay would be longer than expected, thus allowing the extra data to be meaningful in the time-delay analysis. Figure 2.19 presents the 2m2 light curves of HE0047-1756 shifted in magnitude for visual purposes.

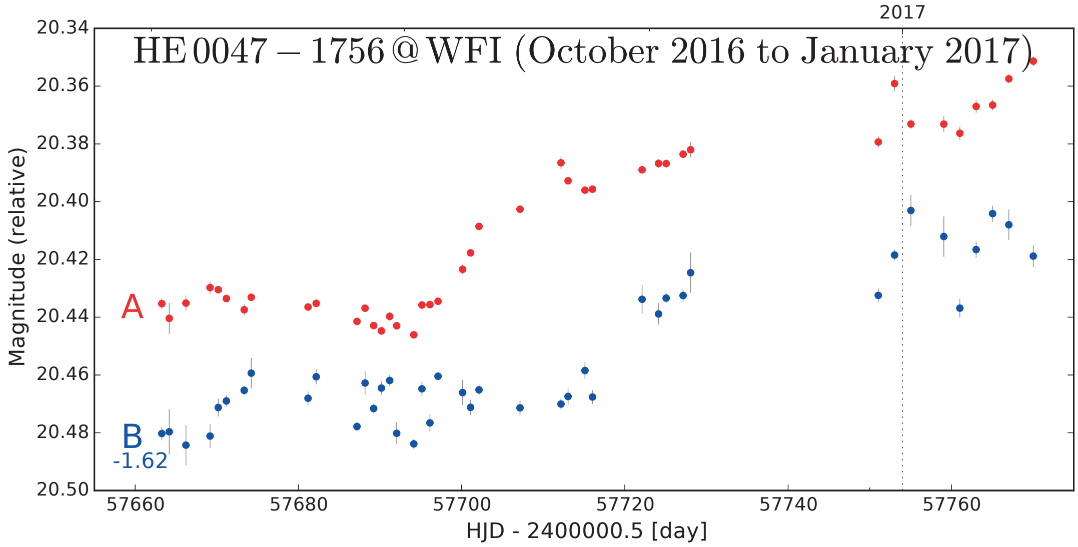


Figure 2.19: 3.5-month WFI light curves of the lensed quasar HE0047-1756. The B light curve has been shifted in magnitude by an offset of -1.62 mag, for visual purposes.

We can note that the typical precision of the WFI measurements is of a few mili-magnitude, which was indeed expected. The scatter of the measurements seems to be sometimes larger than that. This is for example strikingly visible in the B light curve, in the chunk of observations taken in January 2017 (Modified Heliocentric Julian Days (MHJD) from 57750 to 57770). The origin of the scatter is hard to precisely pinpoint; it depends of course on the various steps and calibrations performed during the data reduction that have not been explicitly included in the computation of the photometric error bars, but also on the way the intrinsic and extrinsic variability of the lensed images is modeled (see Sec. 3.2).

Probably the most interesting thing we can learn from Fig. 2.19 is that the high-cadence sampling and improved photometric precision allows us to see a structure in both curves, a plateau followed by an increase of luminosity at MHJD~57700 in A and MHJD~57715 in B. Such a small-scale feature would have been barely notice in the Euler data. Observing such a feature is already sufficient to "discover" a plausible time delay, although observing many of them that are compatible between them is necessary for a precise and accurate determination of the time delay. For that reason, the monitoring of HE0047-1756 at the 2m2 started again in June 2017, this time for a whole 8-month season.

2.5.4 DES J0408-5354

DES J0408-5354 (Lin et al., 2017) is among the first strongly lensed system that have been discovered by the DES collaboration (α(2000): 04h08m21.72s; δ(2000): -53°53'59.5"). It consists of a central galaxy surrounded by three blue lensed quasar images with magnitude ~ 20, and an additional redder and fainter image that is either alone or seen merged with the image of another galaxy in the line of sight; spectroscopy of the system was not able to discriminate between the two cases. The source and lens redshifts are respectively $z_s = 2.375$ and $z_d = 0.597$. The discovery of DES J0408-5354 in early 2016 made it a perfect new and exciting target for the 2m2 monitoring campaign, although being a bit faint to reach the mili-magnitude precision in photometry. The monitoring campaign with the 2m2 started in October 2016 and lasted six months. In parallel to it, the DES collaborators tried to reproduce with numerical models the spatial configuration of the lens (Agnello et al., 2017). The original DES image of the lens had a poor angular resolution and showed no extended features to strongly constrain the modeling. However, by using the best images taken with WFI after a few weeks of monitoring only, I produced a deconvolved image of the whole system of much better quality than the DES original shapshot. Figure 2.20 presents a side-by-side comparison of the DES and deconvolved WFI cut-outs as well as the Gemini i-band acquisition image used to position the slits in order to get the spectrum of the images and lens galaxy. We can notably distinguish in the WFI cut-out part of an Einstein ring around the lens galaxy, as well as three external blobs of unknown origins circled in blue.

Although the lens modeling techniques and principles will be detailed later in this manuscript (see Sec. 4.2.2), I would like to briefly introduce a few general concepts here that are of importance to understand the role of the WFI monitoring in the present context. Using the lensing equations (see Sec. 2.1.1), a proper numerical model of the lens must be able to reproduce the position of the lens galaxy and observed images, their magnification ratios, the presence of extended structures (arcs or rings) and the measured time delays, if any. To do so, it might require to explicitly consider in the modeling the presence of other deflectors of smaller mass around the lens galaxy that act as *secondary* lens. If no time delays are available, the lens modeling can have a very useful predictive power. We saw earlier that equation 2.11 links the lens models, the time-delay distance and the time delays, and is used in time-delay cosmography to infer a value for the cosmological parameters folded in the time-delay distance. But in the case where no time delays are measured, the equation can be used the other way around; by assuming a "standard" cosmology[7], one can roughly predict the expected time delays from the lens modeling.

---

[7] Typically a flat universe with H_0 =70 km s^{-1} Mpc^{-1} and $\Omega_m = 0.3$

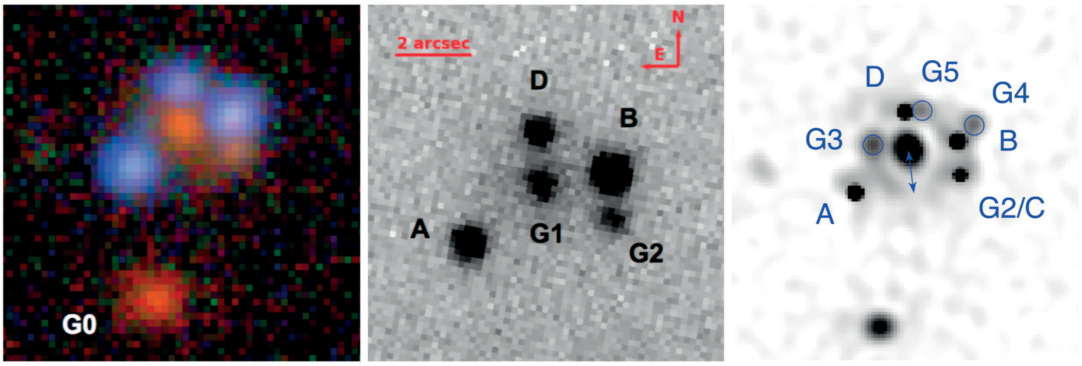


Figure 2.20: Cut-outs of the strongly lensed system DES J0408-5354 with different instruments. From left to right, discovery DES *gri* color image, Gemini *i*-band acquisition image and WFI *r*-band deconvolved 640s exposure. The blue lensed images are labeled A, B and D. The lens galaxy is labeled G1, the perturbing galaxy at the south of the system is labeled G0 and the redder lensed image is labeled G2 or G2/C due to its yet unknown nature - lensed image mixed or not with an additional perturber. Additional blobs of unknown origin that appears on the deconvolved image only are labeled G3, G4 and G5. The blue arrow on the right panel indicates the apparent size of the Einstein ring of ~1.6". Cut-outs adapted from Lin et al. (2017) (left and middle) and Agnello et al. (2017) (right).

Agnello et al. (2017) presents the detailed modeling of DES J0408-5354 using the constraints from the WFI deconvolved exposure. My contribution to this publication was to provide the detailed deconvolved WFI exposure, which is one of the main observational data around which the modeling is done. The soon-to-be-accepted version of the paper is reproduced in Appendix A, and I summarize here the main conclusion in terms of time-delay predictions. Two families of lens models were investigated, one with only one central perturber (the galaxy labeled G1 in the cutouts of Fig. 2.20) and another one with a second smaller perturber at the south-west of the main galaxy, labeled G2. Both models predicts that four lensed images should appear, labeled A, B, C and D. The model with a single perturber predicts that image C should be bluer and brighter than observed. However, the model with two perturbers, providing the second perturber G2 is located very closely to the position of the C image (hence its labeling G2/C) predicts a much more fainter C image in accordance with the observations. Although the evidence from the data would tend to favor the double perturber models, the most obvious way to discriminate between the two solutions at this point is, without surprises, to compare the time-delay predictions to the measured time delays.

The six month monitoring campaign at the 2m2 of DES J0408-5354 was complemented by observations at the Euler telescope, at the average cadence of one point every five nights. The visibility of DES J0408-5354 from La Silla ranging from July to April, the Euler monitoring started three months earlier than the 2m2, in July 2016, for a total of 9 months. The combination of the two data sets is presented in Fig. 2.21. Although the photometric precision of the 2m2 data set is close to the mili-magnitude on the brightest image (B), we can see from the

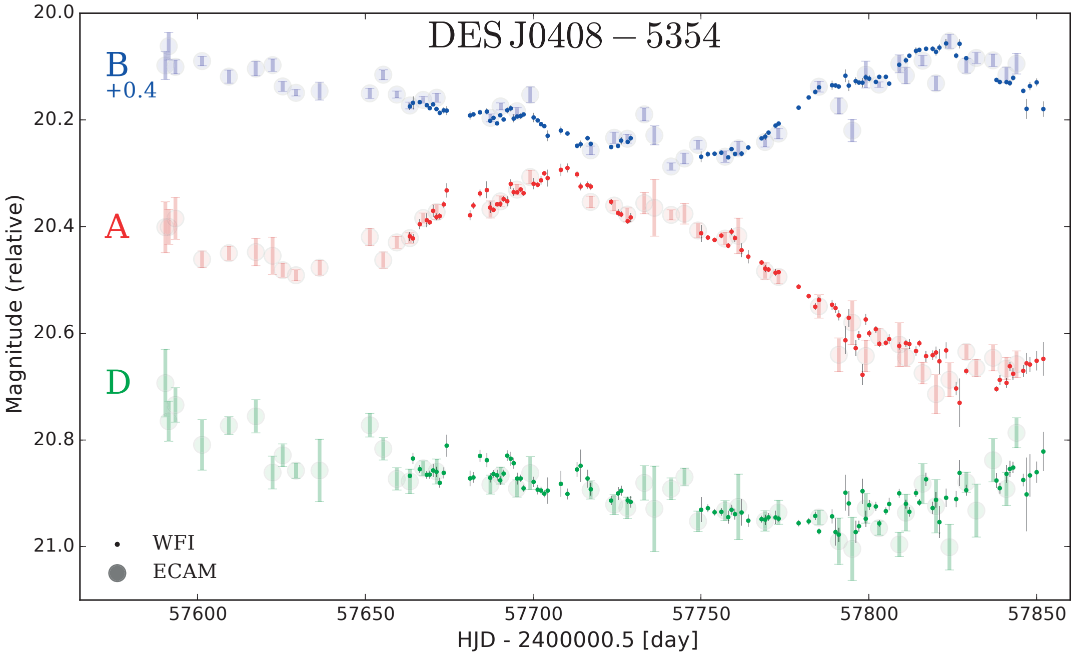


Figure 2.21: 9-month light curves of the lensed quasar DES J0408-5354. The left panel presents the original Euler and 2m2 data with only a shift in magnitude of the B component, for visual purposes. The temporal gap in the 2m2 data around MHJD=57730 corresponds to a ~ 20 day maintenance of the instrument. Adapted from Courbin et al. (2017).

data that it is clearly not the case for the fainter one (D)[8]. Since the latter is around a hundred time less bright, this was of course expected. When exploring the plausible time delays for that system by simply shifting the A, B and D components in time and magnitude by eye, a satisfying solution is that the variations in B occur ~113 days after A. The AB delay is especially well defined, thanks to the peak observed in Fig. 2.21, at MHJD~ 57820 in B and MHJD~ 57710 in A. From these considerations, what Agnello et al. (2017) concludes is that models with free, additional perturber(s) are more likely to reproduce the observed time delay of ~113 days between A and B than models with a single perturber in the center. Nevertheless, observations of additional blobs in the deconvolved 2m2 cut-out (labeled G3, G4 and G5 on the right panel of Fig. 2.20) are currently of unknown nature, and better follow-up imaging and spectroscopy are needed to conclude. The recent published results on DES J0408-5354 are definitely helping in that regard.

2.5.5 PG1115+080

PG1115+080 has been discovered serendipitously by (Weymann et al., 1980), and identified as the first triple-imaged quasar (α(2000): 11h18m17.00s; δ(2000): +07°45'577") at redshift $z_s = 1.722$. It is actually a quad in a fold configuration (Hege et al., 1981), whose two brightest

---

[8] Note that the faintest image (G2/C) is so faint that we do not present any light curve of it here.

images are separated by roughly 0.5 arcseconds only. Although being the second lensed quasar discovered, it received less interest from the community than Q0957+561 in the years following its discovery, certainly because of the difficulty to observe constraining features for time-delay cosmography in this system. The redshift of the lens has been determined in the late 1990's only, as $z_l = 0.311$ by Kundic et al. (1997); Tonry (1998), and identified as being part of a small group of galaxies (Kundic et al., 1997). In the same years, (Impey et al., 1998) observed an Einstein ring around the lens galaxy in the infrared. The lens galaxy proved difficult to observe at first (Henry and Heasley, 1986) and was finally identified as an ellipse (Yoo et al., 2005). Subsequently, precise astrometry of the system was complicated to achieve and had to be refined over the years (Kristian et al., 1993; Courbin et al., 1997).

Regarding the monitoring campaigns of PG1115+080, two independent data sets were published prior to the start of the high-cadence monitoring at the 2m2. Schechter et al. (1997) presented one monitoring season from various telescope at an average weekly cadence taken during the years 1995/1996. A decade later, Vakulik et al. (2009) presented data acquired at a weekly cadence as well at the Maidanak telescope, covering three observing seasons from 2004 to 2006. These data sets gave raise to various time-delay estimates, not always in good agreement with each other (see Sec. 3.4.4). Figure 2.22 presents the light curve obtained after a full season of monitoring at the 2m2 telescope with WFI. Similarly to the two previous 2m2 lenses, we can distinguish a few unmistakable features in the light curves, especially a well-marked dip around MHJD=57800 in A that appears a few days later in B. The last two months of monitoring, however, display a much poorer sampling and photometric precision due to the terrible weather affecting La Silla Observatory[9].

2.5.6 Future monitoring strategies from simulations

In October 2017, the 2m2 will achieve its first year of monitoring for COSMOGRAIL. The light curves of DES J0408-5354 and PG1115+060 already obtained brilliantly demonstrate that the program can successfully achieve its main goal, i.e. detect small enough, well-sampled features in the light curves in order to determine precisely and accurately the time delays. However, there are still pending questions about the best strategy for the year two and beyond. Namely, one can wonder about the necessity of pursuing the monitoring of some targets for a second season. The goal of a monitoring campaign is to measure time delays as accurately and precisely as possible, yet ultimately the measurement will be used to infer a value of the Hubble constant. In that context, if the gain in precision from another season of monitoring is negligible, it is then be better to observe another new target instead and increase the sample of usable strong lenses in time-delay cosmography.

In the scope of high-cadence monitoring, this question reduces to an apparently simple one: how do the precision and accuracy of the time-delay measurements evolve according

---

[9] At the time this sentence is written, the dome of 2m2 telescope is reportedly covered by a ~30 centimetre-thick smooth layer of snow.

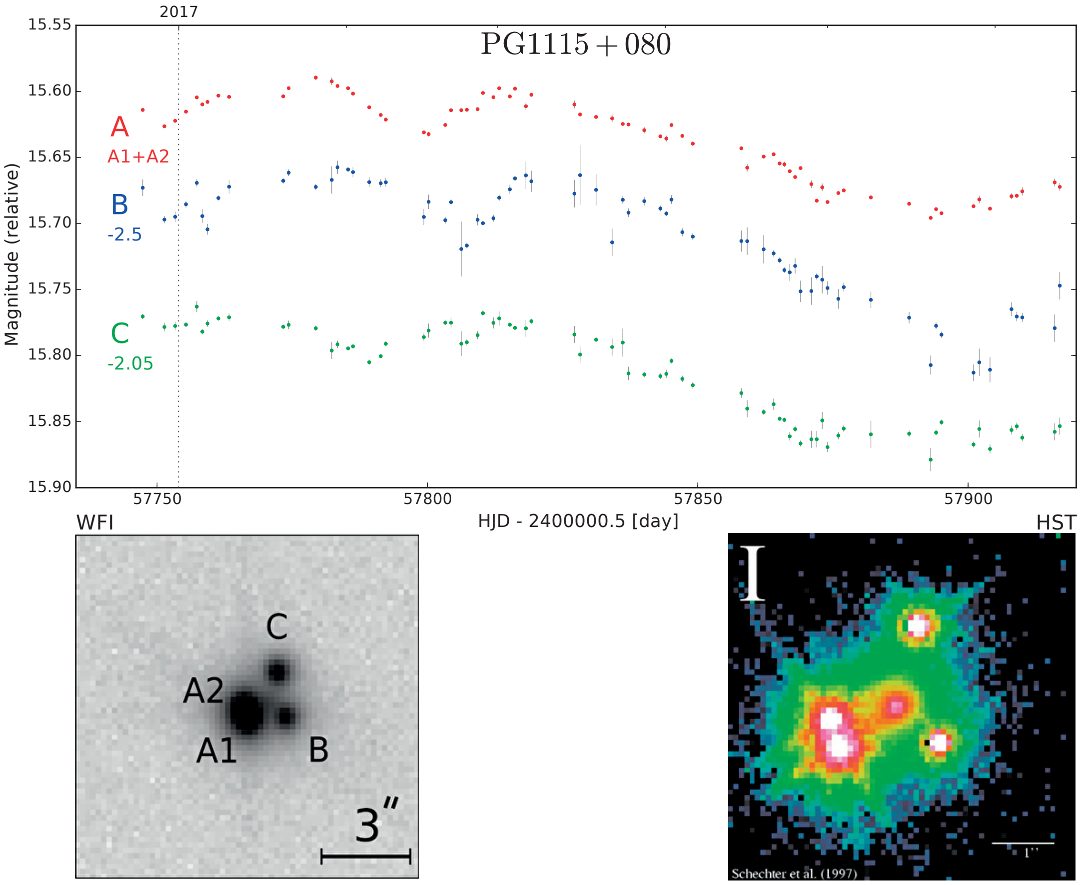


Figure 2.22: 1-year WFI light curves of the lensed quasar PG1115+080, shifted in magnitude for visual purposes. The bottom panels show a single 300-seconds WFI exposure of the quasar (left) and an I-band HST exposure adapted from Kochanek et al. (1998) (right).

to the observational parameters such as the monitoring cadence, exposure time, sources brightness, quasar variability and duration of the monitoring campaign. In order to answer that questions, we performed numerical simulations based on the decade-long light curves of the COSMOGRAIL program. The idea is to create a large set of simulated light curves that respect the properties of the already monitored lensed quasars, add short-scale variability representing the high-frequency and small amplitude signal and see how our state-of-the-art curve shifting techniques perform on these mock data. The two critical aspects of such a work are i) how to generate the simulated light curves and ii) how to robustly assess the quality of the time-delay estimates from these curves. These two questions are addressed in detail in a forthcoming publication (Courbin and Bonvin, prep). I present here the main ideas behind the two points mentioned above and some of the conclusions we can draw from the results.

Generating mock light curves requires us to understand what composes a real set of observed light curves. Putting aside the observational constraints for the moment (observing cadence, photometric precision, correlated noise, etc...), we follow the ideas presented in Tewes et al. (2013a) and break our light curves down into the following components:

- A smooth intrinsic variability curve common to all quasar images. Such variations occurring in the timescale of months to years can be accurately modeled from the long baseline of COSMOGRAIL data.
- Smooth extrinsic variability curves, individual to each lensed image. These curves represent the assumed "slow" microlensing variations that randomly affect the individual images. Again, they can be accurately modeled from the long baseline of COSMOGRAIL data.
- Fast intrinsic variability of small amplitude, that occurs in the timescale of days to weeks (Mushotzky et al., 2011) and affect similarly all quasar images. These variations are by definition not visible in the COSMOGRAIL data.

In order to build our mock light curves, we need to properly mimic these three components. A key ingredient in our modeling is the use of *free-knot splines*. The next chapter dedicated to the time-delay measurement techniques presents the free-knot splines in greater details (see Sec. 3.2.1); it is, however, not necessary to precisely understand how splines work to understand what is presented in the following.

The central assumption we make is that the two smooth components that mimic medium to long-term intrinsic and extrinsic variations are of lesser importance relative to the short-term variations introduced by the fast intrinsic variability. The latter is precisely the component we want to test our sensitivity to, as observing short-scale variations is the key to a fast, precise and accurate time-delay determination. In practice, we use in the simulations the smooth intrinsic and extrinsic variations modeled on the COSMOGRAIL data of HE0435-1223 - although in principle, any quasar with a long baseline of observations would be suitable as well. Since our simulations are expected to be much shorter than the 13-years-long observations of HE0435-1223, we can pick the starting date of the mock light curves randomly in the range covered by the observations, and thus marginalize our results over a large variety of smooth intrinsic and extrinsic variabilities generated form the same initial data.

To mimic the fast intrinsic variability, we chose to follow once again the information we get from the COSMOGRAIL data set. Assuming that the time delays measured in Bonvin et al. (2017) are correct, we shift the four observed light curves of HE0435-1223 in time and magnitude accordingly and stack them into a single curve. We then compute the *moving average* of the stack, such that the resulting light curve has the same sampling than the simulations. The assumption here is that if the fast intrinsic variability is not visible in the individual light curves, stacking them after being correctly shifted in time should increase the sensitivity up to the point that a high-frequency signal should appear. We then compute the stacked light curve *power spectrum*, that will serve as the reference to be reproduced in the simulations. To do so, we tweak the initial smooth intrinsic variability signal used to generate the mock light curves by adding power-law noise in controlled frequency windows so that its power spectrum matches the reference.

Once the intrinsic and extrinsic variability curves have been set, one yet have to sample the mock light curves from them. This implies notably to chose a sampling cadence, seasonal and total duration of the monitoring campaign, an average photometric precision as well as associated noise of the observations, and of course time delays that we want to retrieve. These *simulation parameters* being only partially constrained by the observing conditions, we have a certain freedom in choosing them and thus to explore their impact on the overall precision and accuracy of the time-delay measurements. Table 2.3 presents a detailed view of all the parameters of the generative model. In practice, for a given set of simulation parameters, we generate hundreds of simulated light curves from our generative model whose duration spans the 13 years of the original ECAM data. Then, we extract in each set of mock curves a sample of the duration of our choice. We apply to each sample a *point optimizer*, i.e. an algorithm that yields an estimate of the most probable time delays. Since we know the "true" time delays that were set during the construction of the mock curves, we can compare the results of the point optimizer to the truth and draw all kind of statistics from the results.

Figure 2.23 presents the typical predictive plots produced by the analysis. All the simulations parameters are fixed, except for the one that is explored, indicated in the x-axis of each plot. Each diamond synthesizes the results of a specific point optimizer applied to 200 simulations, marginalizing over the slow intrinsic and extrinsic variations. The y-axis indicates the median relative error of the point estimator for each batch of 200 estimates, in percent: it relates to the true delay put in the simulations, indicated in the title of each panel. The color of each diamond relates to the percentage of estimates that have an individual relative error larger than 5%. Finally, the size of the diamonds represents the standard deviation of the estimates; it is to a certain extent redundant with the median relative error, hence the absence of nominal scale in the plots. Without surprises, we see a positive correlation between the desired behaviours (small median relative error, standard deviation of the errors and fraction of outliers, i.e. estimates with a relative error > 5%) and the campaign duration, photometric precision and frequency of the observations.

The interest of such plots is to predict what are the observational requirements in order to reach a certain threshold in accuracy. We can see notably that extending the monitoring over one year is only marginally helpful in reducing the median relative error and fraction of outliers. However, the gain between a 3-4 months and a full season (8 months) is much more distinct. Similarly, increasing the exposure time in order to reach a mili-magnitude precision is only decisive if the deconvolution noise does not becomes dominant. This means that improvements in the photometric precision must go along with the data reduction and night-to-night calibration accuracy. The most decisive factor of all is certainly the sampling; keeping a frequency of one observation per night is definitely the best way to improve the overall accuracy of the measured time delays even if the monitoring lasts for eight months.

Table 2.3: Exhaustive list of the simulations parameters of the high-cadence monitoring simulations generative model. A short description of each parameter is provided, along with its arbitrary value.

Name	Symbol	Fiducial value	Description
Light curves	–	HE0435-1223 (Bonvin et al., 2017)	Set of light curves used to model the slow intrinsic and extrinsic variations.
Intrinsic spline knot step	η_{int}	35	Initial knot step used to model the slow intrinsic variations of the data.
Extrinsic splines knot step	η_{ml}	300	Initial knot step used to model the slow extrinsic variations of the data.
Duration	N_{days}	250	Duration of the monitoring campaign, including breaks.
Starting date	D_{start}	None	Starting date of the simulated light curve. Random if not defined.
Sampling	freq	2 days	Frequency of the observations.
Sampling randomization	σ_{freq}	$\frac{\mathrm{freq}}{5.0}$ days	Uncertainty on the frequency of the observations.
Photometric error	σ	0.005 mag.	Standard deviation of the photometric error on the fainter image.
Reduction error	σ_{reduc}	0.005 mag.	Standard deviation of the night-to-night calibration error.
Lost nights	–	20%	Fraction of nights lost due to bad weather, facilities issues, etc...
Seasonal gaps	$\mathrm{sea}_{\mathrm{freq}}$ $\mathrm{sea}_{\mathrm{cut}}$	360 days 120 days	Absolute frequency and duration of the seasonal gaps.
Maintenance gaps	$\mathrm{mnt}_{\mathrm{freq}}$ $\mathrm{mnt}_{\mathrm{cut}}$	100 days 8 days	Average frequency and duration of the maintenance gaps.

Let us conclude by emphasizing that such plots predict only the accuracy of the measurements performed by a chosen point optimizer. In real-life cases, since the delays are unknown, the accuracy cannot be measured. Instead, the curve-shifting techniques need to provide an estimation of the precision of their time-delay estimates. Ideally, the precision and accuracy of

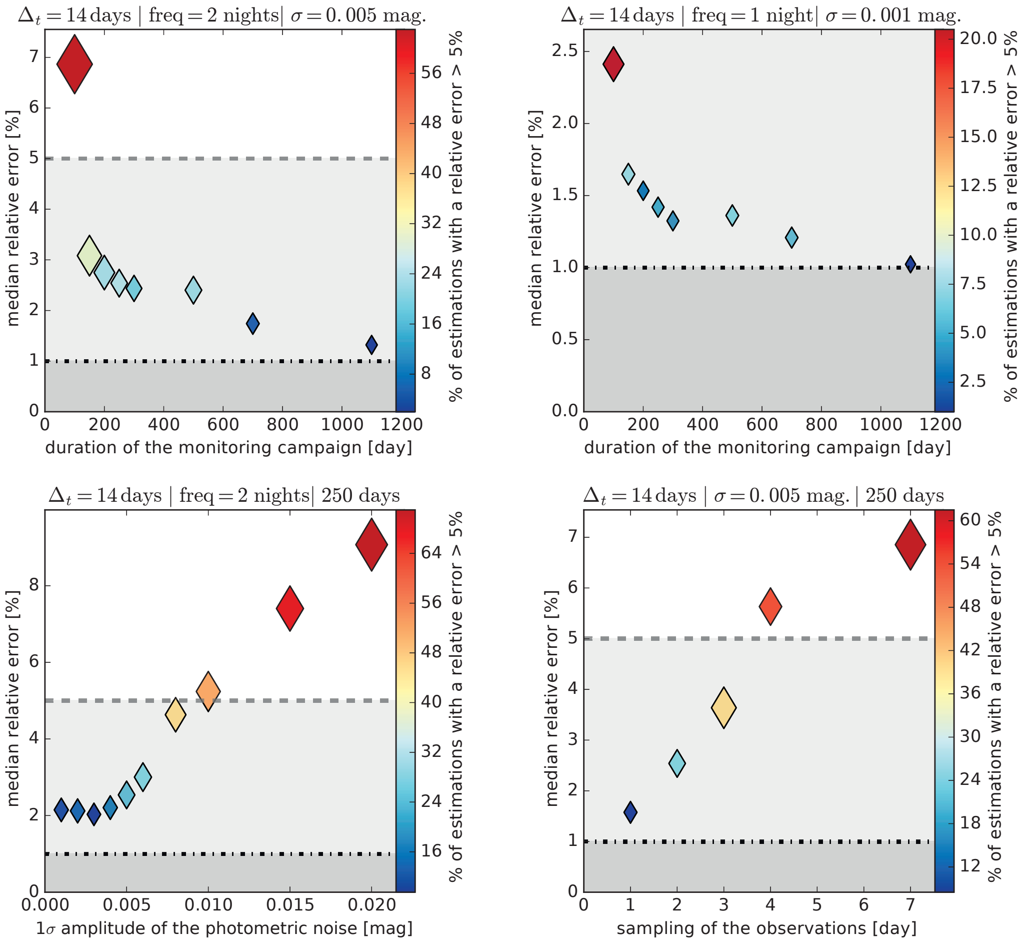


Figure 2.23: Evolution of the time-delay measurements accuracy with various parameters of the generative model. The top row presents the evolution as a function of the duration of the monitoring campaign. The top-left panel uses the fiducial values for the photometric errors and frequency, whereas the top-right panel use more optimistic values, indicated at the top of each panel. The bottom left and bottom right panels present the evolution as a function of the photometric error and sampling of the observations, respectively. Each diamond represents 200 sets of mock light curves created from the same generative model. For each set, the median relative error is indicated in the y-axis, the percentage of estimates with a relative error > 5% is color-coded and the standard deviation is linked to the size of each diamond (the larger the standard deviation, the bigger the diamond). The size of the diamonds is absolute, which allows a direct visual comparison from panel to panel. The light gray and regular gray regions indicates the 5% and 1% limits of the median relative error regions, respectively. The point optimizer used to analyze the simulations is the free-knot spline estimator (see Sec. 3.2.1).

the measurements should be similar, although there is no way to assess it directly on the data.

3 Time-delay measurements

We saw in the previous chapter how to design monitoring campaigns of lensed quasars and reduce the so-obtained data into light curves. In the present chapter, I address the step that naturally follows: I detail how to best analyze such light curves to obtain precise and accurate time-delay estimates that can be used, among other things, in time-delay cosmography.

I start by briefly introducing the general principles of time-delay measurement, lingering notably on the importance of microlensing. Then, I get in the heart of the matter by focusing on the curve-shifting techniques used to measure the time delays. I briefly recall the first attempts at designing efficient curve-shifting techniques before describing in greater details the techniques and formalism I used and developed during the course of my PhD thesis. They are collected in a public python package called `PyCS` (standing for Python Curve Shifting), originally introduced in Tewes et al. (2013a), to which several improvements were brought since its initial release. Major improvements to `PyCS` were notably developed in the scope of the *Time Delay Challenge* (hereafter TDC, Dobler et al., 2015). The goal of the TDC is to estimate the efficiency of state-of-the-art curve-shifting techniques when applied on a large sample of simulated light curves mimicking the output of future large-sky surveys such as the LSST. Having taken part in the TDC, I describe its results and how they can be used to assess the abilities of `PyCS` and forecast the output of future surveys (Liao et al., 2015; Bonvin et al., 2016). Finally, I present the time-delay measurements obtained when effectively applying `PyCS` on a few selected lensed quasars monitored by COSMOGRAIL.

3.1 General principles

Measuring a time delay is very easy to define: given at least two light curves, what temporal shift in time is necessary so that the light curves exhibit the same magnitude variations at the same time ? If the problem is straightforward to solve for two continuous curves, real life presents a few extra issues that make the problem more tricky to solve. First, the signal from the lensed quasar images are unevenly recorded, and according to the Nyquist-Shannon theorem it is not possible to detect a feature that lasts roughly less than twice the monitoring cadence;

put differently, you cannot fit a Gaussian with less than three points. Second, each *epoch* (i.e. the measurements obtained during the course of one observing night) is contaminated by various sources of noise, such as photometric shotnoise, partially correlated calibration and deconvolution noise or flux sharing between multiple lensed images. Third, each light curve can be individually contaminated by long-term highly correlated noise, also called microlensing. The former two issues can be mitigated using bigger telescope with a better spatial resolution at a higher monitoring cadence, although they will still be present at some level in the light curves. The latter issue, on the other hand, depends only on the good will of Nature and is usually the most complicated to address.

3.1.1 A word about microlensing

The microlensing that affect our light curves is the gravitational lensing magnification caused by stars or potentially dark, compact objects moving in the lens galaxy (see Sec. 1.1.5 for a brief introduction). In the usual galaxy-quasar lensing configuration that we are dealing with, these individual microlenses are not massive enough to split the background image into many lensed images like the lens galaxy does. However, on the contrary to strong lensing, the microlenses are moving sufficiently fast so that the resulting microlensing events can be considered as *transient* events, i.e they could be observed from start to finish in a reasonable amount of time. Microlenses in the lens galaxy either magnify of demagnify the already lensed images, thus provoking an independent flickering of the luminosity in each light curve. The time scale of that additional flickering ranges from a month to years and can drastically affect the light curves. For example, the lensed quasar J0158-4325 whose light curves were originally presented in Morgan et al. (2012) and updated in Fig. 2.12 with five extra seasons is strongly affected. Figure 3.1 zooms on three monitoring seasons, from 2005 to 2008. The strong variations visible in the B component are not visible in A, although the delay is likely expected to be shorter than the duration of a single season.

Microlensing in lensed quasars is in itself a very valuable tool (see e.g. Wambsganss, 2006; Kochanek et al., 2007, for a review, as well as some examples listed in Sec. 1.2.4). However, in the context of time-delay measurements, it becomes a real nuisance. By adding long-term correlated signal independently on each individual light curves, the shape of the measured signal is affected in an unpredictable way. Even if we assume that microlensing follows some well-defined statistical properties, we have no clue about the specific realization that effectively impact a given monitoring campaign. And even if the data do not apparently exhibit any evidence of presence of microlensing, assuming there is none could dramatically change the measured time delays.

Our current way of dealing with microlensing follows a completely data-driven approach. We model the *slow* microlensing signal, i.e. the one happening in months to year timescales, directly from the data. So doing, our sole initial assumption is that there are no fast microlensing event. Providing this assumption is incorrect, we could still adapt our smooth microlensing

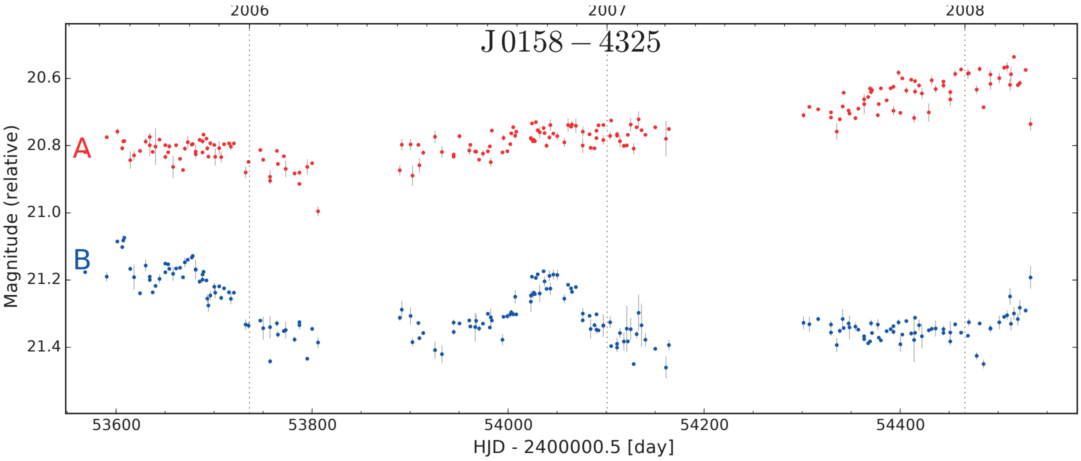


Figure 3.1: Zoom on the late 2005 to early 2008 monitoring data from Euler and SMARTS of the lensed quasar J0158-4325. The peak in the first season of B, around MHJD=53670, as well as the peak in the second season of B, around MHJD=54050 are not visible in the A light curve, although the structures in both light curves post-2011 (see Fig. 2.12) hint towards a time delay shorted than the average season duration.

model so that it matches the observed flickering, up to a certain extent. Other approaches are of course possible, like Kochanek (2004) who explores through brute computational force all the plausible configurations of stellar masses distributions, positions and velocities in order to reproduce the observed microlensing. Although not very subtle, such an approach proved to be very efficient (e.g. Kochanek et al., 2006; Chartas et al., 2009; Dai et al., 2010; Morgan et al., 2010).

Another nuisance induced by microlensing is that, under a few specific assumptions on the nature of the quasar, microlensing can add a systematic contribution to the time delays that can change over long periods of time. The effect is that it introduces a bias of the order of a day on the time-delay measurements, depending on the period during which the monitoring is carried over (Tie and Kochanek, 2017). There are currently no known ways to mitigate the effect of this systematic bias, if not for monitoring the quasars over a sufficiently long period of time so that this microlensing-induced time delay, typically lasting as long as single microlensing events, is marginalized over.

The examples of microlensing events perturbing the time-delay estimates are numerous: microlensing likely played a perturbing role in the controversy on the time-delay of the first lensed quasar Q0957+561 (Pelt et al., 1998), and can completely prevent the measurement of a time delay in the worst-case scenario (e.g. Morgan et al., 2006). It is also one of the reason why the COSMOGRAIL monitoring campaign average duration tends to exceed the decade so that we can properly disentangle the microlensing from the intrinsic variations. Illustrations of microlensing events modeled from the COSMOGRAIL data can be found in (Courbin et al., 2011; Tewes et al., 2013a; Bonvin et al., 2017; Courbin et al., 2017) and are sometimes directly visible in the light curves, such as in the image A of HE0435-1223 reproduced in Fig. 2.7 in which we see a sudden increase of luminosity around mid-2007 with respect to the B, C and D images.

3.2 Curve-shifting techniques

Today's state-of-the art curve-shifting techniques need to be aware of the possible presence of microlensing in the data. It has, however, not always been the case. If the very first monitoring campaigns of Q09567+561 were of course accompanied by time-delay estimates using various king of curve-shifting techniques (see Sec. 1.2.1), Press et al. (1992a) were among the firsts to establish a formalism to measure time delays in a general way[1]. With the multiplication of lensed quasar discoveries and following monitoring campaigns over the years, more and more techniques were developed to address the issue of time-delay measurement, yet with a variable degree of success (see Hirv et al., 2011; Tewes et al., 2013a, for a review).

The curve-shifting techniques evolved along the increase of precision of the observations, as well as the cadence and total duration of the monitoring campaigns. Notably, Pelt et al. (1996) introduced the *dispersion* technique where the light curves were shifted in time, corrected for the possible presence of microlensing and observed flux ratio between the images and then combined into a single light curve. From that combined light curve, it is then possible to compute a dispersion spectrum (hence the name of the technique) that, roughly speaking, quantifies the variability of the curve on which it is applied. The preferred time delay is thus the one for which the dispersion spectrum of the combined curve is the smallest. Put in really simple words, this technique quantify how flat is the subtraction of the two light curves once shifted in time and corrected from microlensing, and the flatter the better. A great strength of that simple procedure is that it is independent from any assumption on the intrinsic variability of the lensed quasar, since it does not need to explicitly model it. Combined to the great simplicity of use, the dispersion technique became very popular. It notably allowed Kundić et al. (1997) to solve the famous controversy around the time delays of Q0957+561, and was successfully applied on various other lensed systems over the years (e.g. Lovell et al., 1998; Biggs et al., 1999; Fassnacht et al., 2002; Courbin et al., 2011; Eulaers and Magain, 2011, among others).

There are, however, no ideal curve-shifting techniques and despite its popularity, the dispersion technique has been proven not being bias-free (Gil-Merino et al., 2002; Hjorth et al., 2002; Pelt et al., 2002; Tewes et al., 2013a). Other techniques developed in parallel over the years are taking advantage of the growth of computing power and the slow but steady adoption of a robust statistical framework by the astronomical community. We can notably cite the recent works of Hojjati and Linder (2014); Aghamousa and Shafieloo (2015); Tak et al. (2016); Rathna Kumar (2017) developed in the scope of the Strong Lens Time Delay Challenge (hereafter TDC, see Dobler et al., 2015; Liao et al., 2015, presented further below in Sec. 3.3.1) as well as our

---

[1] Ironically, that didn't prevented them to get a wrong estimate in the first place

own techniques, regrouped in the `PyCS` module and described in the following.

the *Python Curve-Shifting* module, or `PyCS`, is the COSMOGRAIL contribution to the expanding microcosm of curve-shifting algorithms. It regroups of course several algorithms but also encompasses a data-driven approach to robustly and conservatively estimate the optimal time delays and associated error bars for all kinds of light curves. The seminal work on `PyCS` is presented in (Tewes et al., 2013a,c). Three curve-shifting techniques were originally developed in `PyCS`, one being broadly inspired by the dispersion technique reviewed above. However, further tests on both real and simulated data sets showed that this technique was at best close to the precision of the other `PyCS` optimizers and at worst significantly biased, hence its gradual phasing-out from the most recent COSMOGRAIL publications.

Before delving into a detailed description of the `PyCS` clockwork, let us establish the terminology that we are going to use. A *curve-shifting technique* is a process that takes light curves as an input and yields a corresponding number of time delays with associated error bars in return. The first components of curve-shifting techniques are *point estimators*; they are algorithms whose goal is simply to estimate which time delays fit best the input data. Point estimators can yield other things in addition to the time delays, for example best-fit models of the intrinsic and extrinsic variability. Since point estimators usually minimize a given mathematical quantity by optimizing the parameters a numerical model, they are also referred to as *point optimizers*, or simply as *estimators* or *optimizer*. In the following, these four terms are used indistinctly. The second component of curve-shifting technique are *confidence estimate procedures*. As their name suggests, they aim at quantifying the *confidence* of an associated point estimator on a given data set.

The two curve-shifting techniques presented in the following are called the *free-knot splines* technique and *regression difference* technique. They are both based on point optimizers similarly named, and share the same confidence estimate procedure, also described further down. `PyCS` is still under development, some of which are presented at the end of this section. The code has been made publicly available on the COSMOGRAIL website[2].

3.2.1 PyCS estimator: free-knot splines

The free-knot splines estimator makes use of *B-splines* of degree 3, i.e. a piecewise combination of polynomial functions of degree 3 that are connected to each other by *knots*, where the functions and their first and second derivative are ensured to be continuous. The position of the knots can be freely adjusted on both axis, hence the *free knots* denomination. For a given data set and a fixed number of knots, optimizing the position of the knots to yield the best possible fit cannot be straightforwardly solved (see Molinari et al., 2004, and references therein). More precisely, the fact that the knots are free makes the optimization problem non-linear, where favored solutions tend to see the knots clustering at the same position. In such cases, the continuity of the splines and their derivatives is not ensured anymore which can be

---

[2] http://cosmograil.org/software

problematic regarding the underlying physical assumptions of the phenomenon we want to fit, in our case the luminosity variations of a quasar over time. An efficient approach to tackle this issue has been proposed by Molinari et al. (2004), who introduce as an additional constraint a minimal distance between the knots. This so-called *bounded optimal knots* algorithm for splines fitting, or BOK splines, used in an iterative way, is at the heart of the `PyCS` free-knot splines implementation.

The free-knot splines optimizer takes as an input one or more light curves, and returns a single curve representing the *intrinsic variations* of the lensed quasar. It can also optionally yield individual microlensing curves associated to each input light curves, that we refer to as *extrinsic variations.* Such curves model smooth variations that should be subtracted to the observations so that they match the modeled intrinsic variations of the lensed quasar. Drawing the intrinsic and extrinsic curves requires to decide of a (series of) time shift(s) to be applied to the data, in order for the intrinsic and extrinsic variations to fit at best the data. Finding the optimal time shift(s) is done in an iterative way. At each step, intrinsic and extrinsic splines as well as time shifts are adapted in order to minimize a *cost function* (Eq. 2 of Tewes et al. (2013a), reproduced here for convenience):

$$\chi^2 = \sum_{i=1}^{n} \sum_{j=1}^{N_i} \frac{[m_{ij} - s(t_{ij} + \tau_i) - \mu_i(t_{ij})]^2}{\sigma_{ij}^2}. \tag{3.1}$$

The i index runs over the n light curves, each one composed of N_i data points of coordinates t_{ij} in time and $m_{ij} \pm \sigma_{ij}$ in magnitude. The χ^2 value is estimated at each step for the given time shift(s) τ_i, intrinsic spline s and microlensing splines μ_i, which are all in turn optimized after each iteration. Minimizing the χ^2 value thus yields time delay(s) as well as intrinsic and extrinsic splines.

This optimization problem heavily depends on the *model parameters.* Namely, each spline fitted through the BOK algorithm requires an initial estimate of the time delay(s) as well as an initial *knot step* - the initial separation between two consecutive knots - which controls the adaptability of the spline to fit the data. The choice of an adequate knot step is crucial to ensure a proper behavior of the optimizer, whose sensitivity to the model parameters also greatly depend on the initial data set. In that regard, several diagnostic tools and robustness checks have been developed in `PyCS`, and will be described in Sec. 3.2.3. The effect of the choice of knot steps is, however, directly visible when looking at the light curves.

Figure 3.2 presents the spline fits on the light curves of the quad HE0435-1223, with the data points shifted in time, magnitude and corrected from the modeled extrinsic variability overplotted. The knot steps used are of 35 days for the intrinsic spline and 150 days for the extrinsic splines. Such values are motivated by the resulting fit itself: the intrinsic spline should properly fit the visible features in the stacked data, yet must avoid the "fake" features induced by the

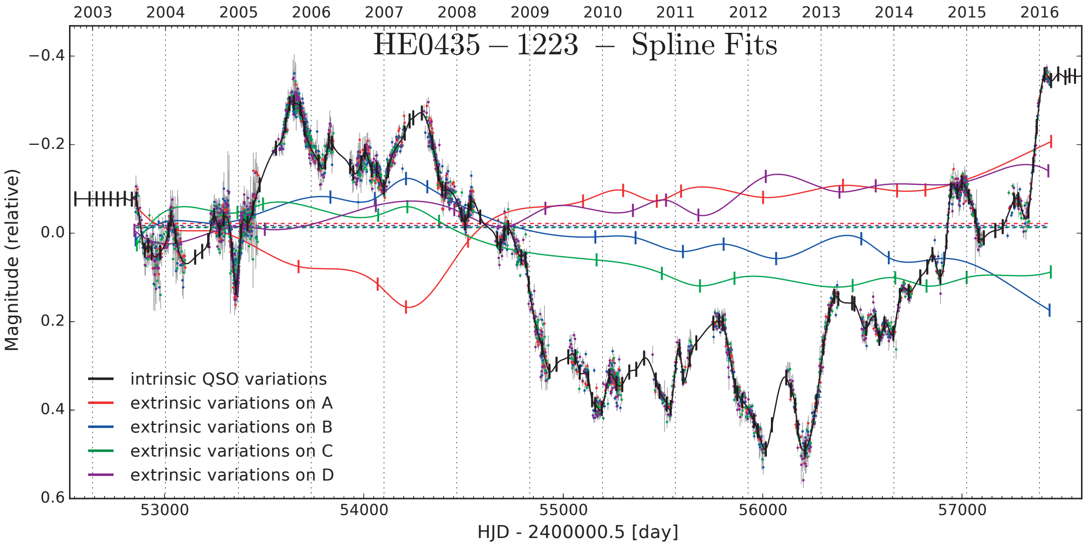


Figure 3.2: Spline fits from the `PyCS` free-knot splines estimator applied to the lensed quasar HE0435-1223, whose light curves were already presented in Fig. 2.7. The intrinsic variability of the quasar, plotted as a solid black line is modeled with an initial knot step of 35 days. Over-plotted are all the data points shifted in time, magnitude and corrected from the extrinsic variability. The four extrinsic variability curves are plotted as solid colored lines and use an initial knot step of 150 days. The optimized position of the knots can be seen as vertical ticks on the splines. Figure adapted from Bonvin et al. (2017).

noise in the data. Such a behavior is commonly referred to as "over-fitting" and can strongly affect the resulting optimization, creating many local minima in the χ^2 minimization. The knot step of the extrinsic splines is chosen so that these spline can only model slow microlensing events, assumed to last at least a few months (see Sec. 3.1.1).

Since its initial development in Tewes et al. (2013a), the free-knot splines optimizer has been used in each COSMOGRAIL publication so far (Tewes et al., 2013b; Eulaers et al., 2013; Rathna Kumar et al., 2013; Bonvin et al., 2017) as well as in the Time Delay Challenge (Liao et al., 2015; Bonvin et al., 2016) where it proved to be among the most robust and precise state-of-the-art point estimator.

3.2.2 PyCS estimator: Regression difference

The regression difference optimizer minimizes the difference between *regressions* drawn from the data. The regressions and associated variances are drawn using Gaussian processes applied on each light curves individually. Gaussian processes are commonly used in astronomy to model time series (e.g. Uttley et al., 2005; Kelly et al., 2014) and can be seen as the generalization of the Gaussian probability distribution to the function space (Rasmussen and Williams, 2006). To perform a regression, the Gaussian process needs to know how to interpret the data points. Namely, the magnitudes of the data points are not purely random but correlated, and the observed correlation can be reproduced using various forms of correlation functions. There is *a priori* no favored choice of correlation function, yet picking one instead of another can significantly change the output of the Gaussian process regression. Thus, the choice of a correlation function as well as the other numerical parameters associated to it must be carefully explored when performing a regression, similarly to the knot steps used in the free-knot splines optimizer presented above.

In practice (and skipping a lot of mathematical details), the regression on a given set of data points performed by a Gaussian process yields a finely sampled curve interpolating the data with an associated error envelope. Figure 3.3 presents the result of a Gaussian process regression on the light curve of image A of HE0435-1223. One can see that the width of the envelope around the regression, representing the 1-σ variance, widens drastically in the regions without data points, especially between two monitoring seasons. This reflects the expected inability of the regression model to predict what extra observations would have looked like if taken in between two seasons. Similarly to the free-knot splines optimizer, one can visually assess if the regression caught all the expected "real" variations of the intrinsic quasar luminosity variations without over-fitting the noise.

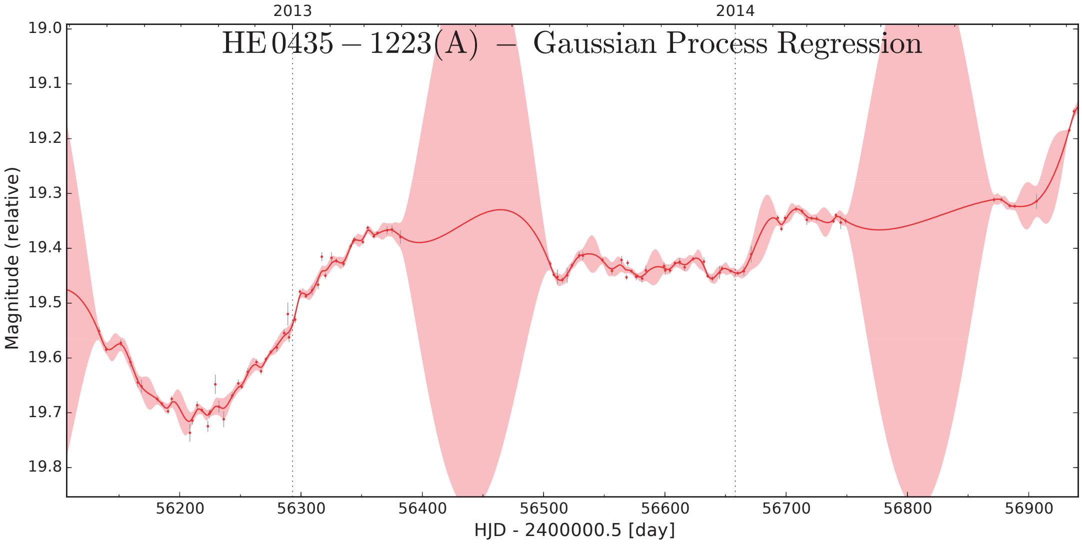


Figure 3.3: Gaussian process regression from the `PyCS` regression difference estimator applied to the image A of the lensed quasar HE0435-1223, for one particular choice of covariance function (Matern kernel) and associated parameters (amplitude parameter of 2.0 mag, a scale of 250 days and a smoothness degree $\nu = 1.5$) following the choice presented in Bonvin et al. (2017). The solid red line represents the regression with its associated 1-σ envelope in shaded red. The original data points and error bars are over-plotted.

Once the regressions and associated variances have been drawn on each light curves individually, they are shifted in time and subtracted pair-wise. From each subtraction results a *difference curve*, whose associated variance is simply the quadrature of the individual regressions' variance at each epoch. From the difference curve is then computed the so-called *weighted average variation*, or WAV, that is defined as follows (Equations 3 to 5 of Tewes et al. (2013a), reproduced here for convenience):

$$WAV(f) = \frac{\sum_{j=1}^{N-1} |\hat{f}'(t_j)| w(j)}{\sum_{j=1}^{N-1} w(j)}, \qquad \hat{f}'(t_j) = \frac{f(t_{j+1}) - f(t_j)}{t_{j+1} - t_j}, \qquad w(j) = \frac{2}{\sigma(t_j) + \sigma(t_{j+1})}, \tag{3.2}$$

where the j index runs over the N regularly sampled points of the regression at time t_j, magnitude $f(t_j)$ and variance $\sigma(t_j)$. The WAV defined above is the smallest when the sum of the absolute derivatives of the difference curve $\sum_{j=1}^{N-1} |\hat{f}'(t_j)|$ is the smallest, i.e. when the difference curve is visually flat. This happens when the structures of the two regression curves cancel out. Thus, finding the time delay becomes a single-parameter minimization problem: each initial time shift yields a value for the WAV which, once minimized, informs us in return about the optimal time delay between the pair of light curves considered. We can note that this method does not need any prior assumptions on the contamination by microlensing of the initial data. Thus, a minimal condition for the regression difference optimizer to yield results that can be trusted is that the amount of matching intrinsic features between the two light curves exceeds the individual microlensing signal. Or, put differently, trusting the regression difference technique requires to assume that the variations introduced by microlensing in the difference curve, if any, are not strong enough to significantly bias the WAV minimization. Applied to quads, the regression difference technique yields *independent* pair-wise time-delay estimates, in the sense that the WAV on the various difference curves are minimized independently from each other.

One flaw of the Gaussian process regression estimator, although minor, is that drawing a regression is computationally costly. The classical implementation scales in $\mathcal{O}(N^3)$, N being the number of data points in the original light curve; the `PyCS` implementation that makes use of the `PyMC` module (Patil et al., 2010) uses some shortcuts speeding up the process, yet being still much slower that a spline fit. In a context where it takes decades to acquire enough data points, spending a few extra hours on time-delay measurements is not really problematic. Yet, new implementations of the Gaussian process regression like the one proposed by Foreman-Mackey et al. (2017) that scales in $\mathcal{O}(N)$ could be interesting to consider as well. That being said, a great strength of the regression difference optimizer is that it follows a completely different approach to the curve-shifting problem than the free-knot splines optimizer. Where the free-knot splines can be seen as a *parametric* approach where all the parameters are optimized simultaneously (knots position and time shifts), the regression difference clearly separates the data-fitting part and the time-shifting part. Additionally, the free-knot splines optimizer requires an explicit modeling of the microlensing, where the regression difference does not consider it at all. That makes the two methods ideally complementary to each other. There is *a priori* no reason to prefer one estimator to the other providing they both perform similarly well in terms of precision on the original data. In such cases, having the two techniques yielding either similar or discrepant time-delay estimates is a great way to robustly assess the accuracy of the final results.

3.2.3 PyCS confidence estimation procedure: measuring time delays and estimating their errors

The two point optimizers described above simply determine a time delay between two curves, but do not estimate the error on that measurement. To properly measure time delays, a single application of these point estimators on the data is not enough: we need to come up with a method to estimate for each time delay an *average value* and associate to it a *variance*. The formalism associated to this confidence estimation procedure adopted in `PyCS` follows the work of Tewes et al. (2013a), that is summarized here. In addition are also presented potential improvements to the procedure, that are going to be implemented in future COSMOGRAIL publications.

In the following, we compute mean time-delay values and associated errors for a fixed choice of optimizer and associated method parameters (i.e. knots steps of the free-knot splines estimator or covariance function of the regression difference estimator). What we want to explore is how well these point estimators perform on a given data set, and use various intermediate control procedures to compare which set of method parameters is the most adequate. Figure 3.4 presents a schematic view of the whole `PyCS` confidence estimation procedure, that visually synthesize what is described in the following.

1) Mean time delays and intrinsic variance

The first step is to estimate the *mean* time delay(s) yielded by the point estimator when applied repeatedly to the original data, but from a different *starting point*. Indeed, the `PyCS` point estimators need an initial estimate of the time delay(s) on which the optimization is build. Depending on what this starting point is, the optimizers could either converge towards their absolute minima or end up in another local minima. The WAV of the regression difference optimizer depends only on a single parameter - the time shift - but the χ^2 of the free-knot splines optimizer depends on many more parameters, like all the knots 2-d position. In both cases, the minimization tries to find the minimum on the parameters hyper-surface, and the influence of the initial time shift(s) must be assessed. In practice, after eyeballing a first few potential guesses, feeding them to the point estimators and looking at the results we can have a rough but solid idea of which initial guess is the most likely. The result of the point estimator from this best initial guess is then used as mean *initial time shifts*, around which other plausible initial estimates are drawn. The range width of these initial estimates, or *starting point interval* is left to the user, and choosing a more or less wide starting point interval should reflect the quality of the data and previous knowledge about the system analyzed; light curves not exhibiting many features should have a broad starting point interval, and

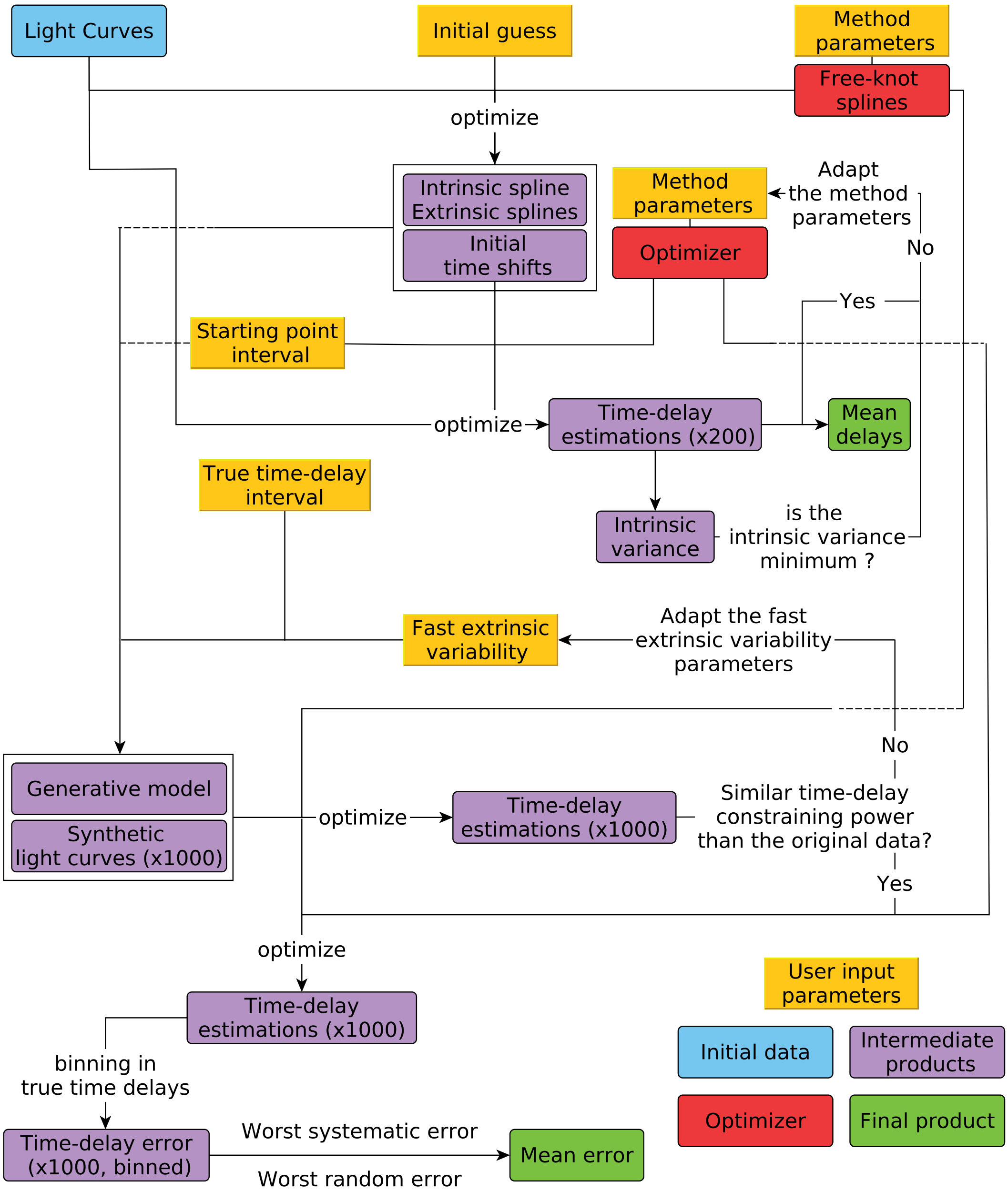


Figure 3.4: Schematic view of the PyCS confidence estimation procedure, as presented in Sec. 3.2.3. The initial data are in blue, the intermediate product in purple and the point optimizers in red. The input parameters whose choice of value are left to the user are in yellow. The final products, i.e. mean time delays and associated error bars are in green.

vice-versa. For example, the starting point interval used in the analysis of the light curves of HE0435-1223 in Bonvin et al. (2017) is broad enough to encompass the 2σ error bar of the previous measurement of the time delays by Courbin et al. (2011).

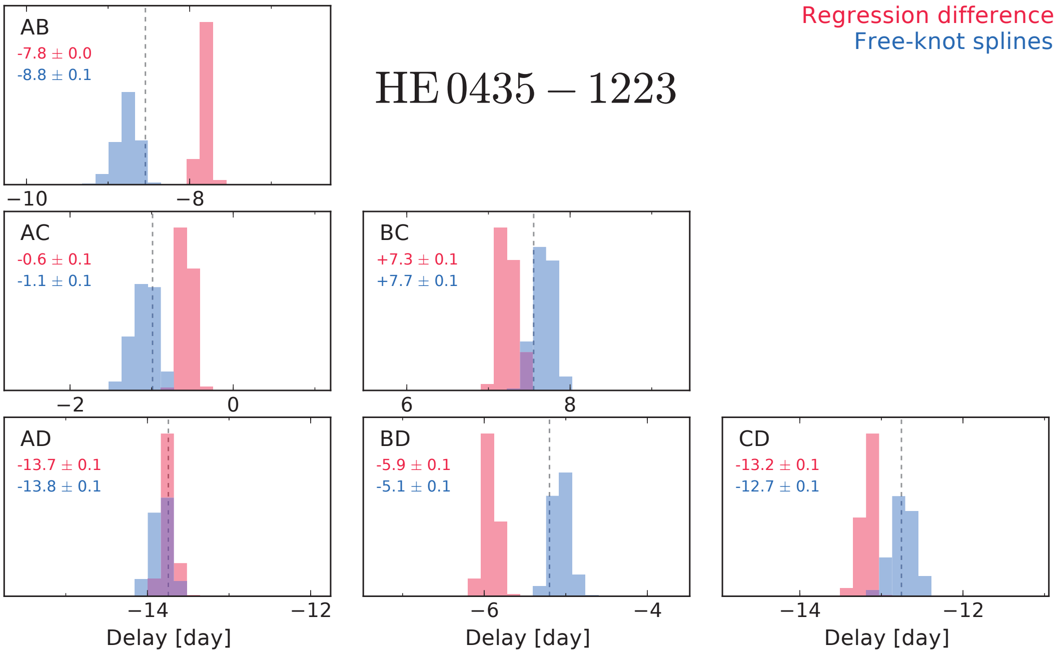


Figure 3.5: Histograms of the intrinsic variance of the point estimators for all the light curves pairs of HE0435-1223. The vertical grey dashed lines represent the initial guessed delays around which the starting point interval is built. For each delay and each estimator, the mean and variance of each histogram are given in the corresponding panels.

Figure 3.5 presents the result of that procedure for both the free-knot splines and regression difference optimizers applied to the light curves of HE0435-1223 presented in Fig. 2.7. The method parameters of each point estimator are the same than the one used in Bonvin et al. (2017). The starting points were randomly chosen in a range of ±6 days around the same initial guess, indicated by a vertical gray dashed line. Each optimizer ran from 200 different starting points, resulting in a distribution of time delays with an associated *mean* and *intrinsic variance*. The means are taken as the final time-delay estimates for each optimizer. However, the intrinsic variances should not be mistaken for the final error on the associated time-delay measurement, that are computed through a more complete procedure detailed further below. Instead, the intrinsic variances are used here as a way to assess that the choice of method parameters do not over-fit the data. If such was the case, the artificial features created in the spline fits or in the regressions would contaminate the intrinsic signal, effectively creating unphysical minima in the optimizers hyper-surface explored during the minimization procedure. As a result, the intrinsic variances would be much larger, i.e. the optimizers would get stuck in one of the many unphysical hyper-surface minima. Thus, a good way to assess that our choice of method parameters is suitable is to make sure the intrinsic variance of the considered optimizer is as small as possible. Let us note that at this stage, two different optimizers that do not agree with each other (i.e. the difference of their means is much larger than the average of their intrinsic variances) like in the AB and BD panels of Fig. 3.5 does *not* signify that both estimates are in tension with each other.

2) Generative model of synthetic light curves

The second step consists of estimating the *uncertainty*, or the effective error affecting the point estimator on a particularly data set. This is done by exploring the estimator sensitivity to the statistical properties of the data. To do so, `PyCS` follows a Monte-Carlo approach where *synthetic light curves* with known *true time delays* are drawn and fed to the point estimators in order to assess both their bias (systematic error) and variance (random error). When building synthetic light curves, special care must be taken to assess that they have the same *time-delay constraining power* than the original data, i.e there should be qualitatively the same amount of features to which the optimizers are sensitive to in the synthetic and in the real light curves.

In order to do so, we construct a *generative model* based on the original data from which we sample our synthetic light curves. The generative model can be divided in two parts. First, the base components that stay the same for any set of light curves generated by the model. These are the *intrinsic quasar variability and smooth (or "slow") extrinsic variability curves* that are created by running the free-knot splines optimizer on the data. Second, the *"fast" extrinsic variability*, or correlated noise added to each light curve individually, and that is randomized for every realization of the generative model. That noise can be seen as the combined effect of fast microlensing, deconvolution noise, etc... that have not been yet explicitly considered in the modeling. In practice, that fast extrinsic variability is modeled through power-law noise (Timmer and Koenig, 1995) whose frequency window, amplitude and power spectrum exponent can be fine-tuned to obtain the desired time-delay constraining power. Note that this formalism has many similarities with the one used in Sec. 2.5.6 to draw simulated light curves for high-cadence monitoring. The main difference is that here, we are not adding any assumed fast intrinsic variability to our synthetic light curves; even though we would have physical evidence that such variability exists, doing so would add too much time-delay constraining power to the synthetic light curves. In other word, possible high-frequency signal in the original data is either detected by the optimizer and modeled in the intrinsic quasar variability, or not modeled at all.

To measure the time-delay constraining power on a set of light curves, we run the free-know splines optimizer on the set and compute the residuals of the fit once corrected by the assumed microlensing (i.e. how far is each epoch from the intrinsic spline fit on Fig. 3.2). We can then compute two statistics from these residuals. First, their standard deviation σ and second, the number of runs r (see Eqs. 10 and 11 of Tewes et al. (2013a)). A *run* is defined as a sequence of adjacent residuals that are either positive or negative, regardless of their individual amplitude. The higher (lower) the number of runs r in a data set, the more correlated (anti-correlated) the residuals are. The hypothesis we make in `PyCS` is that these two statistics define sufficiently well the time-delay constraining power. By making sure that the generated synthetic light curves have the same σ and r than the original data, we assess that the former are suited to estimate the uncertainty of the optimizer. The tricky part of this process is that we do not know *a priori* what the σ and r of the synthetic light curves will be before running the free-knot spline technique on a bunch of them, since these two statistics depend on the residuals of the spline fit. Thus, what we do in practice is to iteratively adjust the parameters of the power-law noise used to model the fast extrinsic variability until the time-delay constraining power of the synthetic light curves corresponds to the observations.

3) Uncertainty and robustness checks

Once the correct parameters to model the fast extrinsic variability have been determined, a large number of synthetic light curves are drawn from the generative model. Similarly to the study of the intrinsic variance, the starting point of the optimizer is also randomly chosen in an uniform *starting point interval* defined around the the initial time shifts. In addition, the known true time delay(s) put in the synthetic light curves are also chosen in an uniform *true time-delay interval* chosen around the guessed delay. The reason to do so is to test if the optimizer is, as expected, sensitive to real features in the light curves or if there is a preferred value to which the optimizer is sensitive to regardless of the true delay - an undesired feature associated to the so-called *lethargy* of the optimizer. (Molinari et al., 2004; Rathna Kumar et al., 2013). Similarly to the starting point interval, the true time-delay interval should reflect our *a priori* knowledge and confidence about the plausible range of time delays that could adequately fit the data. In practice, the width of these two intervals are often the same.

For each set of synthetic light curves, the point optimizer yields an estimate of the time delay(s), than can thus be compared to the value of the true time delay(s) to assess the overall accuracy of the optimizer. We illustrate this step by producing for each pair of light curves a scatter plot, binned in true time-delay value that allows us to examine the general behavior of the optimizer. Such a plot is reproduced in Fig. 3.6, again for the lensed quasar HE0435-1223.

We compute the bias and variance of all the estimates in each bin, that we associate respectively to the systematic and random error of the corresponding point estimators. The behavior of the systematic error (the shaded rods on Fig. 3.6) must be closely monitored, as it inform us on the lethargy of the optimizer. A lethargic optimizer would see its bias decreasing with increasing true time delays, thus visually producing a "descending staircase". No such effect is visible in the present case. The final uncertainty on the time-delay measurement(s) is computed as the combination in quadrature of the worst bias and the worst variance among all the bins. This is of special importance if we suspect the estimator to be somehow biased as it allows us to take into account the worst possible scenario in a conservative way, at least over the range of true time delays that have been explored.

Once these final time-delay estimates have been computed, the study of the intrinsic variance and point estimator uncertainty is repeated using different method parameters for the optimizers. This also require to recreate the generative model used to draw the synthetic light curves. Comparing the resulting time-delay estimates allows to assess the general robustness of the whole process. Typically, if a small change in the method parameters produces completely different results, it is a sign that the optimizer's results cannot be fully trusted. Similarly, in the cases where the baseline of observations is long enough - typically over a decade - the original

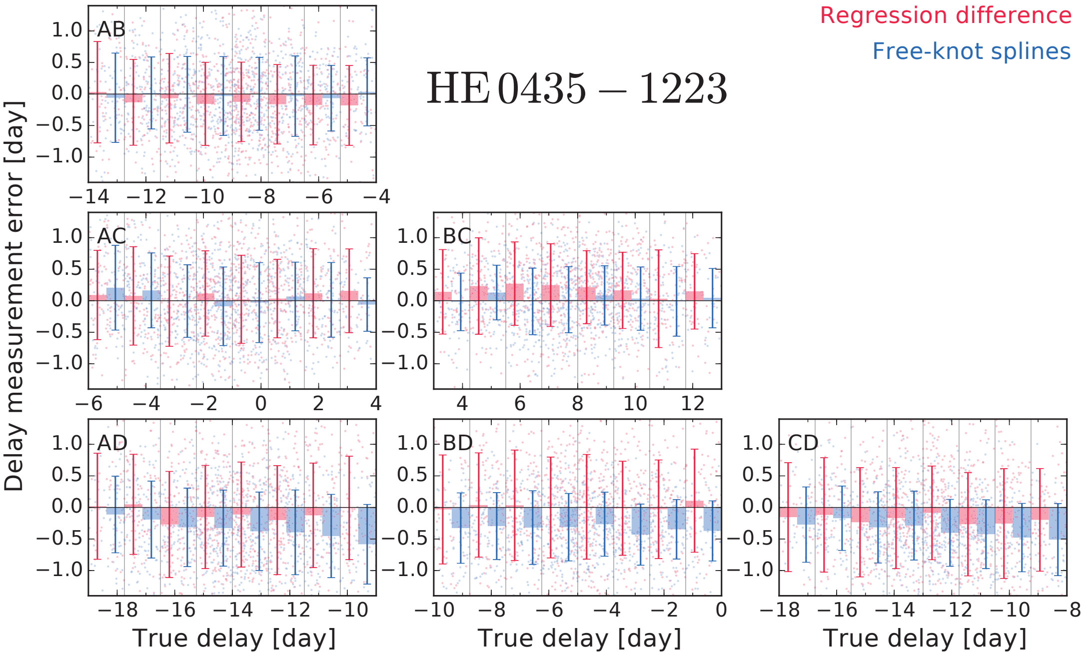


Figure 3.6: Error of the free-knot splines and regression difference optimizers on a set of 1000 synthetic light curves drawn from a generative model of HE0435-1223. The x-axis represents the true time delays put in the synthetic light curves that are split in 8 equivalent bins separated by thin gray solid lines. The small points represents the individual estimates. The shaded rod and associated vertical error bar represent the mean and standard deviation, respectively, of all the individual estimates falling in the corresponding true time-delay bin.

data set can be split in two or three chunks of a few years of data each and the time delays estimated independently on each chunk. Figure 3.7 presents the final time-delay estimates of HE0435-1223 after 13 years of monitoring, as presented in Bonvin et al. (2017). It compares the estimates resulting from the free-knot splines and regression difference techniques as well as the results published in Courbin et al. (2011) that used a technique inspired by the dispersion method of (Pelt et al., 1996). In addition, the data have been split in three chunks of 4+ years each, on which the time delays have been similarly measured. The figure shows that in the present case, there is no evident sign of bias or over/underestimation of the time-delay uncertainties, thus supporting the results obtained on the full data set.

3.2.4 PyCS: possible improvements to the current confidence estimation process

The whole confidence estimation procedure described above combined to the state-of-the-art curve-shifting optimizers used in `PyCS` allow us to to provide both precise and accurate time-delay measurements, while keeping a conservative approach. The performances of `PyCS` have been studied in the scope of the Time Delay Challenge (Dobler et al., 2015; Liao et al., 2015) and in further details in Bonvin et al. (2016) and will be discussed later on. Yet, before moving on this part, I would like first to present two possible ways to develop the confidence

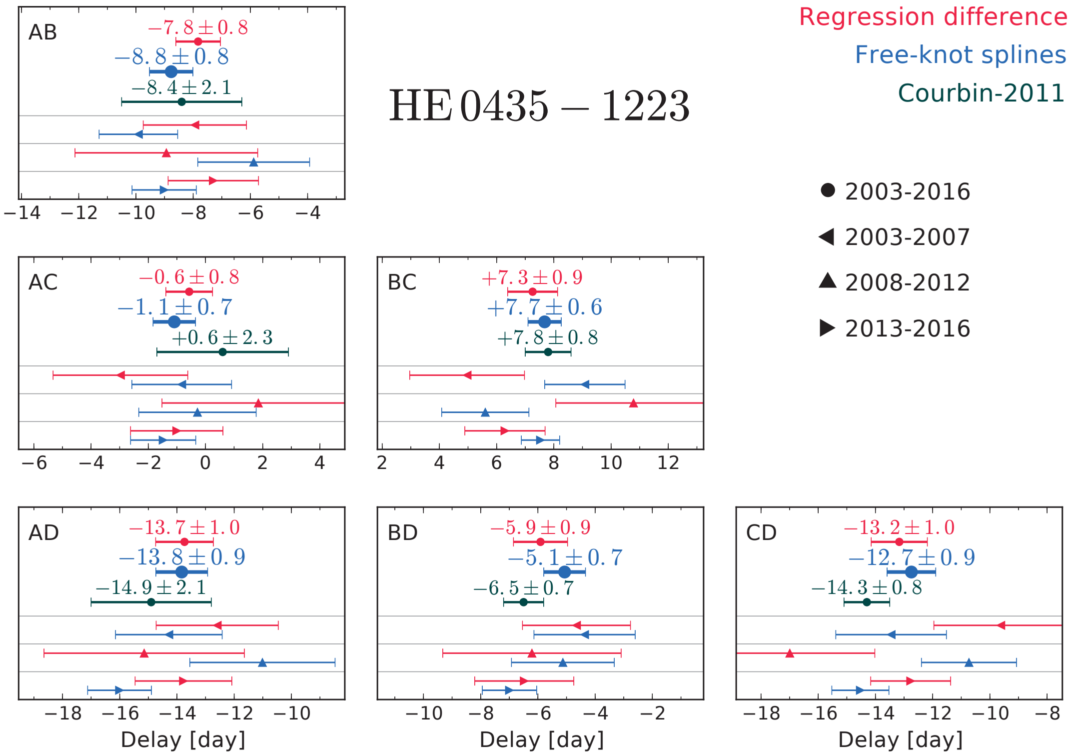


Figure 3.7: Time delays for the six pairs of quasar images of HE0435-1223. The upper part of each panel displays the time-delay measurements and associated 1σ error bars for the free-knot splines and regression difference techniques as well as the results from Courbin et al. (2011). The lower part of each panels displays the the time-delay measurement carried out individually on three chunks of 4+ years each. Adapted from Bonvin et al. (2017).

estimation procedure of `PyCS` further.

1) The generative model

The first idea concerns the generative model of synthetic light curves. Currently, it makes use of the free-knot splines optimizer and requires an explicit model of the quasar intrinsic variation as well as the slow extrinsic variability associated to each light curve. Yet, we might want to consider another kind of generative model that does not explicitly requires slow extrinsic variability components. Such would be useful for example when it is not clear from the data if extrinsic variability is present or not. Although the current generative model can work with no extrinsic variability, it still requires the `PyCS` user to make the initial assumption that there is none. The improvement would thus be to draw each synthetic light curve directly from the corresponding observed light curve.

Doing so can be hazardous, since it does not take into account the physical assumption that there is a common pattern in all the light curves. Whether this is important in the drawing process or not is yet to be properly assessed. However, what we know for sure to be of importance is to assess that the synthetic light curves have the same time-delay constraining power than the data. In order to measure it, the generative model needs to make use of an optimizer that either interpolate or extrapolate a *smooth model* of the individual light-curve variations from the data since such a model is necessary to compute the scatter of the residuals σ and number of runs r used to quantify the time-delay constraining power. In that regard, both an individual free-knot splines fit or a Gaussian-process regression could do the trick. For each real light curve, an independent smooth model of the variability would be produced and serve as a basis on which fast extrinsic variability is added. The time-delay constraining power would also be measured independently on each synthetic light curve, and the parameters of the extrinsic variability would be iteratively adjusted.

The development of this new generative model is currently in its early phase and no concrete results can be presented yet. However, it is important to understand that this new procedure does not aim to replace the current one, but to provide an alternative that can be used in parallel. Similarly to the two point estimators currently used in `PyCS`, having two different generative models allows us to cross-check the results against the presence of potential sources of biases of unknown nature that would not have been detected otherwise.

2) Time-delay likelihood and covariance

The current way of expressing time-delay measurements for quads is to give six dependent and consistent time-delay estimates, without any estimation of the covariance coefficients between these six measurements. We only check visually that the correlations between residuals behave as expected, as explained in Sec. 3 of Tewes et al. (2013a). The motivation for giving these six measurements is that we do not want to pick *a priori* one of the quasar images as reference without any strong motivation for it. Delivering these six *dependent* estimates does contain part of the information that would otherwise go into a covariance matrix associated with giving only three "independent" delays. Indeed, if we label the lensed images A, B, C and D, a tight covariance between the AB and AC time-delay measurements, for example, would tell that the possible values of the time delay BC are already somehow constrained. Currently, we just give these time delays AB, AC and BC without any extra information on how they relate to each other. When fitting a lens model using the time delays as constraints (see Sec. 4.2.2), one currently selects *a posteriori* three "independent" delays with small error bars.

Yet, assuming that these three "independent" estimates are truly independent is wrong: the true time-delays are, but not their measurements. Furthermore, selecting three time-delays among six is arbitrary and using all six time delays sounds certainly redundant. To solve this, the idea is to propose a classical form of the time-delay likelihood as the standard output of `PyCS`. By making the assumption that the distribution of the error on each individual time-delay measurement follows a Gaussian distribution - the less constraining assumption following the principle of maximum entropy that is reasonably well verified on our synthetic light curves - the time-delay likelihood can be written as follows:

$$p(\Delta t_{\mathrm{model}}|\Delta t_{\mathrm{obs}}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\Delta t_{\mathrm{model}} - \Delta t_{\mathrm{obs}})^T \Sigma^{-1} (\Delta t_{\mathrm{model}} - \Delta t_{\mathrm{obs}})\right). \quad (3.3)$$

Here, $\Delta t_{\mathrm{model}}$ are the time delays for which we want to compute a probability density function, Δt_{obs} are the time-delay measurements on the sample of synthetic light curves, n is the number of pair of images ($n = 6$ for a quad) and Σ is an estimation of the covariance matrix of Δt_{obs}.

Writing the likelihood using three *independent* time-delay measurements and the associated $3\,x\,3$ covariance matrix should in fact capture the full information contained in our six dependent measurements, even if the reference light curve chosen is of poor quality and another choice would apparently seems more appropriate. Note that computing the likelihood using the six dependent time delays is mathematically not allowed since the $6\,x\,6$ associated covariance matrix is by construction singular. To the best of my knowledge, there is no way to modify the equation above to properly take into account that redundancy in the six delays. Thus, we instead provide the four $3\,x\,3$ sub-matrices of time delays and covariance coefficients to compute the time-delay likelihoods and recommend either to pick a reference image, or (better!) to run whatever analysis using the so-constructed time-delay likelihoods as input parameters by picking in turn each of the four images as reference and average the results. Running the analysis with all the possible independent combinations of time delays would be the optimal solution but probably a bit of an overkill.

To be consistent with the conservative approach used when estimating the time-delay uncertainty, the computation of the covariance matrix coefficients should follow the same principles used so far in `PyCS`: as explained in Sec. 3.2.3, we draw simulated light curves from a generative model, we give them "true" time delays randomly selected in a chosen range of plausible time delays around our initial estimate, run the optimizer on each simulation, bin the results according to their true time delays and then measure the systematic and random error in each bins. We then combine the worst systematic and worst random error to get the final time-delay error estimation. When constructing the covariance matrix, we still keep this procedure for the diagonal coefficients that correspond to the error on each time-delay measurement used hitherto. To compute the off-diagonal coefficients, we do a two-dimensional binning of the true time delays (since we compute the covariance of two different time delays), using potentially larger bins to ensure that there are enough estimates per bin, and compute the covariance coefficient in each bin. This time, however, we cannot simply pick the highest or lowest coefficients among all the bins to build the likelihood, because doing so will not ensure the resulting likelihood to be the most conservative one. The two options available at this stage are:

1. consider all the possible combination of off-diagonal coefficients ($n\,x\,n$ 2d-bins would

mean n^6 combinations for a quad) and take the less constraining likelihood;

2. compute a single likelihood using all the samples (i.e. without binning) when computing the off-diagonals coefficients, and control that when doing the 2d-binning there are no large variations between the coefficients computed from one bin to another.

The first option, although technically more conservative than the second one is more complicated to implement; not only the number of possible likelihoods becomes significantly large for $n > 3$ but defining which likelihood is the less conservative can be tricky if their shape do not follow a well-known distribution. Thus, the second option is at the moment favored, although it still requires extensive testing to assess that such an approach makes fully sense. Let us note that in order to get a scatter from bin to bin not dominated by random fluctuations, we ought to draw more simulated light curves from the generative model than before. 10'000 seems to be a reasonable number from the testing performed so far.

Figure 3.8 illustrates the process described above applied on 10'000 simulated light curves created from the generative model of HE0435-1223. The time-delay errors and mean covariance coefficients are indicated on top of each panel. Experiments about the use of these covariance coefficients are currently conducted inside the H0LiCOW collaboration and, if conclusive, will be implemented in future publications.

3.3 Forecasts from time-delay measurements

The upcoming years will see the emergence of wide area imaging surveys that will deliver an increasingly number of interesting targets for cosmology, as detailed in the introduction of Sec. 2.5. Among them, the LSST is predicted to discover thousands of new lensed quasars (Oguri and Marshall, 2010; Collett, 2015) with the particularity of repeatedly observing the whole survey area every few days and thus directly providing light curves. In that context, it is crucial to forecast what could be the cosmological output from this wealth of data. Such predictions would also allow to finely tailor the LSST baseline strategy, where small changes could potentially significantly improve the cosmological outcome.

3.3.1 The Time-Delay Challenge

The Time-Delay Challenge (hereafter TDC) is a blind challenge developed by LSST collaborators. Its goals are twofold: first, allow the scientific community working on time-delay measurements to test the robustness of their curve-shifting techniques and determine what kind of precision and accuracy they are currently able to achieve. Second, possibly adapt the LSST baseline strategy by adapting the cadence, duration, etc... of the observations depending on the results of the TDC. In practice, the TDC has been planned in various steps of increasing difficulty. Prior to the first real challenge, labeled TDC1, was the introductory step labeled TDC0. The latter consisted of a simple set of 50 simulated pair of light curves of unknown

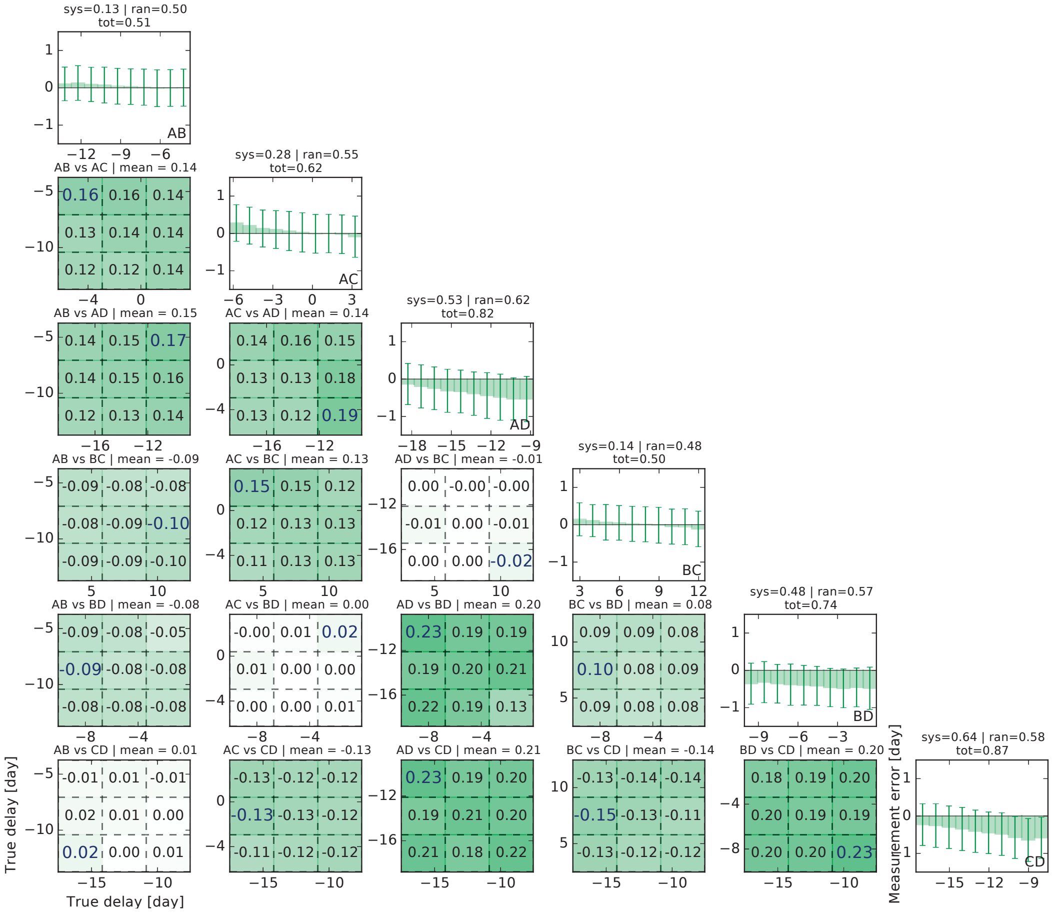


Figure 3.8: Overview of the time-delay error and covariance coefficients computed on 10'000 synthetic light curves created from the generative model of HE0435-1223, using the improved procedure described in Sec. 3.2.4. The panels on the diagonal are equivalent to the ones presented in Fig. 3.6 with the results binned in true time delays. The worst random error and worst systematic error are printed in the title of each panel, as well as their combination in quadrature as the total error. The off-diagonal panels follow the same principle with two-dimensional bins of true time delays. The value over-plotted in each bin corresponds to the covariance coefficient computed using only the simulations in the given bin. The higher absolute value in each panel is displayed in blue to ease the visualization. The corresponding pair of time-delays considered is written at the head of each panel, as well as the mean covariance coefficient using all the available simulations, i.e. without considering any binning.

time delays, most of them unaffected by microlensing and modeled as if observed through a single R-band filter. The goal of this introductory step was to make a first selection among the teams interested to participate: the applicants had to send their best guess for the time delays and associated 1σ error for all the 50 pairs of light curves; only the submissions passing an arbitrary quality threshold (detailed below) were thus invited to participate to the TDC1. TDC0 submissions were open one month prior the start of the TDC1 and lasted until its end. Each participating team was allowed to send as many submissions as desired, even if notified that their previous submission failed to pass the thresholds. By doing so, the TDC0 was more of a gateway indicating if a team's curve-shifting technique will be able to perform reasonably well in the real challenge, or not. Nevertheless, on the fourteen teams that participated in the TDC0, only seven qualified for the TDC1. The results of the TDC0 as well as the experimental design of the challenge are presented in Dobler et al. (2015).

The TDC1 was the continuity of the TDC0 but on a much larger scale. It consisted of ~5000 pairs of simulated light curves. Roughly 70% of them were modeled from doubles, where the remaining 30% were modeled from quads and grouped in pairs. The challenge lasted six months, during which the teams could submit only one submission per curve-shifting technique. No feedback on the submissions was provided until the end of the challenge. The format of the submissions, similarly to the TDC0, consisted of the teams providing their best guess for the time delay and associated 1σ error, for a maximal number of pairs. In that sense, the TDC1 was solely a curve-shifting challenge, without any cosmology involved. The simulated light curves were split in five *rungs* of approx ~1000 pairs each. Each rung had its own sampling, cadence regularity, individual season duration and total number of seasons, while marginalizing over the photometric and correlated noise, intrinsic quasar variability and microlensing. It was revealed to the teams at the end of the challenge that in order to ease the comparison between the different rungs, the same generative model was used in each rung (i.e. each pair had a corresponding equivalent in all the other rungs, with the same true time delay), but with the corresponding pairs of light curves randomly mixed-up in each rung in order to avoid the participants figuring it out. Rung 0 contained light curves similar to the COSMOGRAIL results (i.e. five seasons of eight months with a mean sampling of four days) where rungs 1 to 4 were closer to the predicted LSST output, with shorter seasons of longer cadence. The submission were judged according to four different metrics, reproduced below from Bonvin et al. (2016):

1. The fraction f of submitted time-delay estimates,

   $$f = N_{\text{submit}}/N, \tag{3.4}$$

   where N_{submit} is the number of measured time delays and N is the total number of curves.

2. The mean χ^2 between the measured time delay $\widetilde{\Delta t_i}$ and the true value Δt_i weighted using the *estimated* uncertainties δ_i,

   $$\chi^2 = \frac{1}{N_{\text{submit}}} \sum_i \left(\frac{\widetilde{\Delta t_i} - \Delta t_i}{\delta_i} \right)^2 . \tag{3.5}$$

3. The mean "claimed" precision P of the time-delay measurements, computed from the

estimated uncertainties δ_i,

$$P = \frac{1}{N_{\mathrm{submit}}} \sum_i \left(\frac{\delta_i}{|\Delta t_i|} \right). \tag{3.6}$$

4. The mean accuracy A of the time-delay measurements,

$$A = \frac{1}{N_{\mathrm{submit}}} \sum_i \left(\frac{\widetilde{\Delta t_i} - \Delta t_i}{\Delta t_i} \right). \tag{3.7}$$

These four metrics all have, obviously, an importance on their own. The TDC team had a series of expectations regarding how well the best curve-shifting techniques should score on these metrics. For example, the desired goal regarding the precision and accuracy of the techniques is that the precision P and accuracy A both stay below 3%, resulting in $0.5 < \chi^2 < 1.5$. The failure rate $1 - f$ should be lower than 70% (i.e. $f > 0.3$) for the lens monitoring to make sense in the LSST global observing strategy.

I took part in the challenge along with M. Tewes and F. Courbin as the `PyCS` team. We used the free-knot spline technique (see. 3.2.1), a modified version of the regression difference technique (3.2.2) and a third very simple technique based on eyeballing the data. Various techniques were used by the six other teams, many of which involve the use of Gaussian processes to produce a smooth model of the light curves. Fully Bayesian approaches and dispersion techniques were also used by other groups, ensuring in overall a wide representation of the existing curve-shifting techniques currently applied on real data.

The results of the TDC1 are presented in Liao et al. (2015), that is reproduced in Appendix A. It features a rung-by-rung and metric-by-metric comparison of the teams' performances (see Figs 5 to 9 and Fig. 13), yet without providing any kind of general ranking. As a general observation, the accuracy A and chi-square χ^2 of the best algorithms (including the two main `PyCS` submissions) were far beyond the initial expectations. The requirements in precision P and fraction and submitted estimates f were, however, only marginally fulfilled, at best. Nevertheless, the results of this first step were encouraging; real data would provide more information through multi-band light curves and the direct imaging of the lens, thus bringing additional constraints when measuring the time delays.

The next step, labeled TDC2 is to start in the upcoming months (P.J. Marshall, private comm.) and will feature multi-band observations as well as a Fermat potential associated to each lens. The participating teams will have to submit their best guess of the Hubble constant H_0 directly compiled from all the light curves. By doing so, the TDC2 moves closer to mimicking the real data outcome of the LSST. Furthermore, the TDC2 will also provide a single metric to rank the teams, H_0, easier to relate to for people not well versed in the field of time-delay measurements.

3.3.2 Assessing the `PyCS` abilities

The Time-Delay Challenge with its thousands of simulated light curves was a perfect occasion to test the performances of `PyCS`. This is even more true since the generative model used to draw the synthetic light curves as described in Dobler et al. (2015) is completely independent from `PyCS`. This performance study has been thoroughly performed in Bonvin et al. (2016) reproduced hereafter, and whose key points can be summarized as follows:

1. The analysis is split in two Stages. Stage I aims at *discovering* the time delays through a visual inspection of the light curves. Stage II builds on that initial guess to *refine* the time delay and associated error bars using `PyCS`. Stage I estimates were performed through a java web interface especially developed for this purpose. At least two different human beings have seen each pair of light curves, whose estimates were classified in four different confidence categories: secure, plausible, multimodal and uninformative. Such an approach, although tedious at first sight proved to be extremely valuable in rejecting potential outliers that would not have been detected without visual inspection, as stressed out in (Liao et al., 2015). Combined together, the secure and plausible Stage I estimates consisted on 67% of the total sample of light curves. Less than 1.3% of these were labeled as *catastrophic outliers*, i.e. more than 20 days away from the true time delay. In addition, we also present an automated discovery algorithm developed after the end of the TDC1 whose goal is to speed-up and improve further the confidence of the Stage I results.

2. Stage II used two `PyCS` optimizers: the free-knot splines (Sec. 3.2.1) and a modified version of the regression difference (Sec. 3.2.2) where the smooth modeling of the individual light curves usually performed with Gaussian processes has been replace by spline fits to speed-up the fitting part. In both cases, running the whole confidence estimation procedure as presented in Sec. 3.2.3 would take way too much time, so we simplified the process by i) adding plain white noise instead of fast correlated extrinsic variability to the mock light curves from our generative model, ii) putting always the same true time-delay in these mock light curves and iii) not fine-tuning the white-noise parameters according to the residuals. The results were nevertheless extremely encouraging; for example, on COSMOGRAIL-like data (rung 0), the average precision P of the free-knot spline was below 3% and the accuracy A below 1%. The resulting χ^2 was thus below 1, meaning that we tended to overestimate the error bars. The regression difference technique using splines was slightly less precise and accurate but more consistent in its error estimation process as its χ^2 was approximately of 1.

3. Taking the so adapted free-knot splines technique as our currently best technique, we analyze in more depths the achievability of time-delay measurements according to the properties of the pair of light curves considered, namely their photometric precision and true time delay. Figure 3.9, partly reproduced from Bonvin et al. (2016), presents the evolution of P and χ^2 according to these two properties. It shows that the best

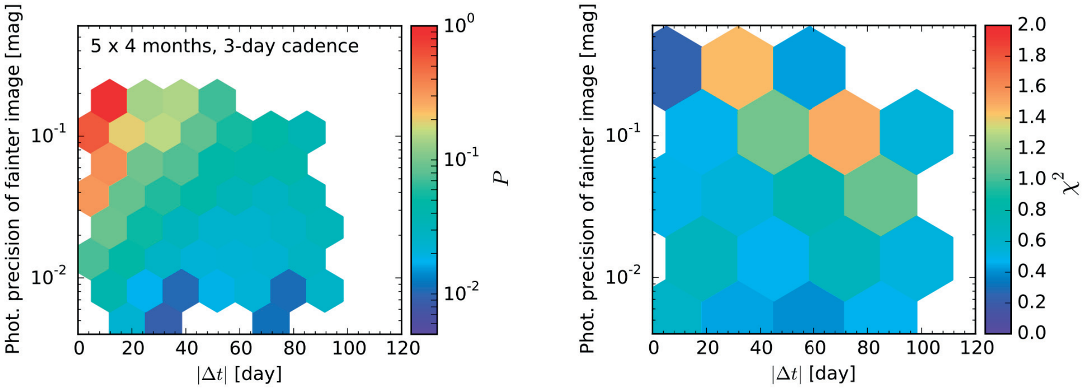


Figure 3.9: Quantitative analysis of the precision P (left) and χ^2 (right) achieved by the free-knot spline technique as a function of the photometric precision of the fainter quasar image and the true time delay. The pairs of light curves used regroup all the secure and plausible Stage I estimates from the rungs 2 and 3 of the TDC1, after rejection of the catastrophic outliers. Adapted from Bonvin et al. (2016).

results are obtained for time delays that are shorter than two-thirds of the season length. Considering the overall performances of the free-knot spline technique compared to the other techniques used in the TDC1, we propose to use our results as forecasts of what can be achieved from state-of-the-art time-delay measurement techniques.

A&A 585, A88 (2016)
DOI: 10.1051/0004-6361/201526704

Astronomy & Astrophysics

COSMOGRAIL: the COSmological MOnitoring of GRAvItational Lenses

XV. Assessing the achievability and precision of time-delay measurements

V. Bonvin[1], M. Tewes[2], F. Courbin[1], T. Kuntzer[1], D. Sluse[3], and G. Meylan[1]

[1] Laboratoire d'astrophysique, École Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland
e-mail: `vivien.bonvin@epfl.ch`
[2] Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany
[3] Institut d'Astrophysique et de Géophysique, Université de Liège, Allée du 6 Août 17, B5c, 4000 Liège, Belgium

Received 9 June 2015 / Accepted 28 September 2015

ABSTRACT

COSMOGRAIL is a long-term photometric monitoring of gravitationally lensed quasars aimed at implementing Refsdal's time-delay method to measure cosmological parameters, in particular H_0. Given the long and well sampled light curves of strongly lensed quasars, time-delay measurements require numerical techniques whose quality must be assessed. To this end, and also in view of future monitoring programs or surveys such as the LSST, a blind signal processing competition named Time Delay Challenge 1 (TDC1) was held in 2014. The aim of the present paper, which is based on the simulated light curves from the TDC1, is double. First, we test the performance of the time-delay measurement techniques currently used in COSMOGRAIL. Second, we analyse the quantity and quality of the harvest of time delays obtained from the TDC1 simulations. To achieve these goals, we first discover time delays through a careful inspection of the light curves via a dedicated visual interface. Our measurement algorithms can then be applied to the data in an automated way. We show that our techniques have no significant biases, and yield adequate uncertainty estimates resulting in reduced χ^2 values between 0.5 and 1.0. We provide estimates for the number and precision of time-delay measurements that can be expected from future time-delay monitoring campaigns as a function of the photometric signal-to-noise ratio and of the true time delay. We make our blind measurements on the TDC1 data publicly available.

Key words. methods: data analysis – gravitational lensing: strong – cosmological parameters

1. Introduction

The methods used to constrain current cosmological models all benefit from independent measurements of the local value of the *Hubble* parameter, H_0 (see e.g. Fig. 48 of Weinberg et al. 2013). One way of achieving a measurement of H_0 is based on time delays in strong gravitational lens systems. The method, first suggested by Refsdal (1964), proposes measuring the differences in the travel time of photons coming from multiple images of a distant source, such as a quasar. This time delay, Δt, is connected to the overall matter distribution responsible for the lensing effect, and to the time-delay distance $D_{\Delta t}$ to the lens, i.e. H_0, with some sensitivity to curvature and dark energy as well (e.g., Suyu et al. 2014).

Exploiting this relationship to constrain H_0 and cosmology in general requires both an accurate mass model for the lens and accurate time delay measurements (see e.g., Suyu et al. 2012; Linder 2011; Moustakas et al. 2009). Modelling the lens mass in an unbiased way is difficult and prone to degeneracies known as the mass-sheet degeneracy (e.g., Schneider & Sluse 2013) and, more generally, the source plane transformation described in Schneider & Sluse (2014). The effect of lens model degeneracies can be mitigated by combining astrometric information from high resolution imaging, measurements of the lens dynamics, priors on the mass density profile of the lens, and an analysis of structures along the line of sight (e.g., Suyu et al. 2014; Greene et al. 2013; Fadely et al. 2010; Treu & Koopmans 2002; Falco et al. 1997; Keeton & Kochanek 1997). Integral field spectroscopy coupled with the adaptive optics that is becoming available on the VLT and at the Keck observatory will be essential in this respect. One of the key ingredient for the method to work at all is, however, the quality of the time-delay measurements, which is the focus of the present work.

In practice, measuring time delays is achievable if the lensed source is photometrically variable. Gravitationally lensed quasars are ideal targets: the quasars can show variability accessible by moderately sized ground-based optical telescopes on timescales of a few days, while the time delays resulting from galaxy-scale lenses are typically of the order of weeks to months (see, e.g., Oguri & Marshall 2010). The intrinsic light curve of the quasar is seen shifted by the relative time delays in the different lensed images. However, this simple situation is often contaminated: microlensing by stars in the lensing galaxy introduces extrinsic variability in the individual light curves with an amplitude sometimes comparable with that of the intrinsic variability of the quasar. To yield competitive cosmological constraints, reliable time-delay measurements with percent-level precision are needed (Treu 2010). An efficient implementation of these measurements has long been hampered by how difficult it is to obtain photometric data for periods of many years at a desirable cadence, which must be close to 1 epoch per few days (Eigenbrod et al. 2005).

COSMOGRAIL is a monitoring program targeting more than 30 strongly lensed quasars using meter-class telescopes, with a cadence of 3 days for the most interesting systems. Recent results include light curves and time-delay measurements that are accurate to within a few percent points in HE 0435−1223 (Courbin et al. 2011), SDSS J1001+5027 (Rathna Kumar et al. 2013) and in RX J1131−1231 (Tewes et al. 2013b). To measure these time delays, we developed and implemented several algorithms in the form of a COSMOGRAIL curve-shifting toolbox named `PyCS`, described in Tewes et al. (2013a).

In the fall of 2013, a first blind time-delay measurement competition named Time Delay Challenge 1 (TDC1) was proposed to the community by Dobler et al. (2015). The two main goals of this open challenge were (1) to homogeneously assess the quality of time-delay measurement algorithms on a common set of realistic synthetic light curves, and (2) to obtain some quantitative informations on the impact of observing strategy (cadence, season length, campaign length) on time-delay measurements. We took part in this challenge and submitted time-delay estimates using the team name PyCS. Liao et al. (2015) give a summary of the results from all TDC1 participating teams, as well as some general conclusions. The present paper is complementary to Liao et al. (2015). It focuses on the `PyCS` methods that we also apply to real light curves, and hence assesses the quality and reliability of the COSMOGRAIL time-delay measurements.

To evaluate our existing techniques with the thousands of light curves of TDC1 under conditions similar to the analysis of COSMOGRAIL data, we separated the problem of time-delay measurement of a light curve pair into two successive stages.

Stage I: we first attempt to discover a plausible time delay, without trying to measuring it precisely. We also evaluated how confident we were that the proposed approximate solution is close to the true time delay and not a catastrophic failure. Owing to the limited length of the monitoring seasons, the randomness of quasar variability, noise and microlensing, this was not possible for every light curve pair of TDC1 or a real monitoring campaign. We note that in the case of TDC1 we had no prior information on the time delay to look for, as we had no knowledge of the mass distribution in the lensing galaxy. Only the light curves themselves were used.

Stage II: for those systems for which Stage I was successful, we then focused on accurately estimating the time delay and associated uncertainty with the `PyCS` techniques, constraining the algorithms to a delay range around the solution from Stage I. As the `PyCS` Stage II methods did not rely on a physical model of the light curves, they would not be able to deal adequately with comparing odds among very different solutions.

This two-stage structure is of general interest beyond `PyCS`, as the stages concern discernible aspects of the time-delay measurement problem. Stage I deals with the quantity and the purity of time-delay measurements, while Stage II deals with their actual accuracy.

The paper is structured as follows. Section 2 summarizes the data from the TDC1 and the metrics used to evaluate techniques. In Sect. 3, we present the approaches we took to address Stage I, while Sect. 4 presents the Stage II algorithms. In Sect. 5 we discuss the results, expanding on the analysis of Liao et al. (2015), and we conclude in Sect. 6.

2. Time Delay Challenge 1

The mock data used in this work are synthetic quasar light curves made publicly available in the context of the TDC1 proposed by Dobler et al. (2015). These data mimic the photometric variations seen in real gravitationally lensed quasars, with different time sampling, number of seasons, and season length. The curves are generated with simple yet plausible noise properties, and include microlensing variability. The dataset is split into five "rungs" or stages that simulate different observational strategies, each rung consisting of 1024 pairs of light curves. The rungs randomly mix microlensing, noise properties, and variability but differ in time sampling, number of seasons, and season length. These differences are listed in Table 1 of Liao et al. (2015, hereafter TDC1 paper).

Participants to the TDC1 were asked to provide their best point estimate $\widetilde{\Delta t_i}$ and associated 1σ uncertainty δ_i for as many pairs of curves as possible. The submitted entries to the challenge were then compared to the true input delays, and evaluated using simple metrics probing different properties of the estimates. The details of how the simulations were set up, as well as a discussion of these metrics are given in Dobler et al. (2015). Results obtained by the blind submissions of the different participating teams are summarized in the TDC1 paper, including our submissions denoted "`PyCS`". For completeness, we briefly summarize the four main metrics:

1. the fraction f of submitted time delay estimates,
$$f = N_{\mathrm{submit}}/N, \tag{1}$$
where N_{submit} is the number of measured time delays and N is the total number of curves.
2. the mean χ^2 between the measured time delay $\widetilde{\Delta t_i}$ and the true value Δt_i weighted using the estimated uncertainties δ_i,
$$\chi^2 = \frac{1}{N_{\mathrm{submit}}} \sum_i \left(\frac{\widetilde{\Delta t_i} - \Delta t_i}{\delta_i} \right)^2. \tag{2}$$
3. the mean "claimed" precision P of the time-delay measurements, computed from the estimated uncertainties δ_i,
$$P = \frac{1}{N_{\mathrm{submit}}} \sum_i \left(\frac{\delta_i}{|\Delta t_i|} \right). \tag{3}$$
4. the mean accuracy A of the time-delay measurements,
$$A = \frac{1}{N_{\mathrm{submit}}} \sum_i \left(\frac{\widetilde{\Delta t_i} - \Delta t_i}{\Delta t_i} \right). \tag{4}$$

To analyse the results in greater detail, we also introduce two modified metrics. First, a "blind" precision,
$$\tilde{P} = \frac{1}{N_{\mathrm{submit}}} \sum_i \tilde{P}_i = \frac{1}{N_{\mathrm{submit}}} \sum_i \left(\frac{\delta_i}{|\widetilde{\Delta t_i}|} \right), \tag{5}$$
where we replace in Eq. (3) the true value of Δt_i by its estimation $\widetilde{\Delta t_i}$. This metric can be computed without knowing of the true time delays; its summation terms are useful, for instance to sort light curve pairs of a real monitoring survey. Second, we introduce a variant of the accuracy
$$A_{\mathrm{abs}} = \frac{1}{N_{\mathrm{submit}}} \sum_i \left(\frac{\widetilde{\Delta t_i} - \Delta t_i}{|\Delta t_i|} \right), \tag{6}$$
where we replace the Δt_i in the denominator of Eq. (4) by its absolute value. While A is sensitive to a bias on the amplitude of $\widetilde{\Delta t_i}$ (i.e., over- or underestimation of delays), A_{abs} responds to signed biases.

Statistical uncertainties on these metrics are computed following

$$X_{\rm err} = \sigma_X / \sqrt{N_{\rm submit}}, \tag{7}$$

where σ_X is the sample standard deviation of the summation terms of $X = \chi^2$, P, and A.

3. Stage I: discovering time delays

To apply any of the PyCS time-delay measurement algorithms (Tewes et al. 2013a) to a pair of light curves, a prior estimate of the time delay is required. Depending on the considered light curves, identifying this delay from the data might be difficult or even impossible. In the following, we describe two approaches to discover rough time-delay estimates (Stage I). Both methods rely solely on the light curves without considering the configuration of the lens system. The first method is based on a visual inspection of the light curves and is the method we used to blindly prepare submissions for the TDC1 (Liao et al. 2015). We developed the second method after the TDC1 results were released. We use the data from the challenge to evaluate the relative merits and drawbacks of each method.

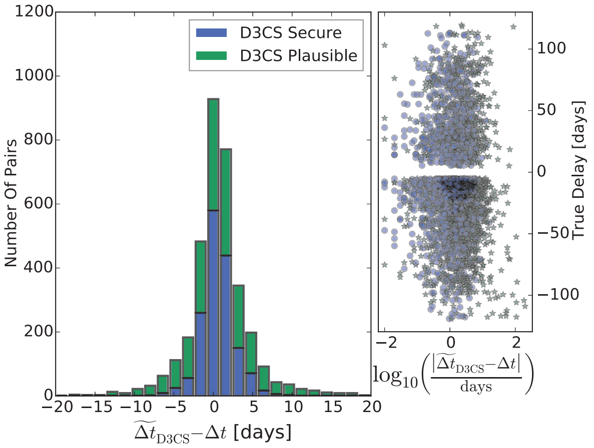


Fig. 1. *Left panel*: stacked histogram of the errors made by the visual time-delay estimation, for the `secure` and `plausible` D3CS samples. All the `secure` estimations are within the displayed range of errors of ±20 days. Only 2.6% of the time delays in the `plausible` sample have an absolute error larger than 20 days. *Right panel*: absolute time-delay error made by D3CS as a function of the true delay for both samples.

3.1. D3CS: D3 visual curve shifting

This method is based on visual inspection of the light curves, in the spirit of citizen science projects (e.g., see the review by Marshall et al. 2015). To ease this process, we developed a dedicated browser-based visualization interface, using the `D3.js` JavaScript library[1] by Bostock et al. (2011). We have made this interface public[2].

The main motivations behind this time-costly yet simple approach were (1) to obtain rough initial estimates for the time delays and their associated uncertainties, and (2) to estimate how confident one can be that the time-delay estimations are not catastrophic outliers. Clearly, visual curve-shifting allows for more freedom than any automated procedure. It also permits dealing in a more flexible way with unusual behaviour of light curves, such as highly variable signal-to-noise from one season to the next, extreme microlensing, or even when the time delays are comparable in length to the visibility period of the object.

Our interface allows users to interactively shift the light curves in time, magnitude, and flux, and to zoom in on sections of the data. It permits the visual estimation of the time delay and of an associated uncertainty. Importantly, the interface also asks to pick one of four choices of confidence category for the proposed solution:

1. `secure`: if a catastrophic error can be excluded with a very high confidence[3];
2. `plausible`: if the solution yields a good fit and no other solutions are seen;
3. `multimodal`: if the proposed solution is only one among two or more possible solutions that are equally satisfactory;
4. `uninformative`: if the data does not allow the estimate of any time delay.

Unlike massive crowdsourcing programs (e.g. `Galaxy Zoo`; Lintott et al. 2008), only four scientists participated in the visual inspection of TDC1 and each pair of curves was measured by at least two independent scientists. The behaviour of different users in terms of time spent per time-delay measurement span a broad range. Fast users spend on average 25 s per object, while slower users spend more than 60 s per object. This includes the time taken to measure a delay, to estimate a 1σ uncertainty, and to allocate one of the four confidence levels described above.

To obtain a single Stage I estimation for each light curve pair, we reduce the results obtained by all four scientists in a very conservative way. We systematically downgrade to `multimodal` the confidence category of pairs with conflicting time-delay estimates.

We define samples combining the four confidence categories as follows: `sec` stands for `secure` only, `secpla` stands for `secure` + `plausible`, and `secplamul` for `secure` + `plausible` + `multimodal`. The combination of all estimations is labelled `all`.

Figure 1 shows the distribution of the error on the time-delay estimation versus the true time delay for the `secure` and `plausible` D3CS subsamples. Table 1 summarizes the results of the D3CS classification and displays the fraction of catastrophic outliers ϵ, i.e. time-delay estimations more than 20 days away from the truth. Notably, the `secure` sample contains 1623 time-delay estimates free of any catastrophic outliers.

Through this simple experiment, we have demonstrated that such an approach is easily manageable for about 5000 light curves. In total the four scientists involved in the visual estimation of the time delays spent 150 h measuring the 5120 delays. We note that 30% of the time delays were measured by three or more users.

3.2. Attempt to design an automated Stage I procedure

Visual inspection of the light curves is a time-consuming process that cannot be repeated many times. Designing an automated method whose efficiency approaches that of D3CS is therefore complementary and would help to minimize the time spent on the visual inspection. We developed such a method after the end

---

[1] Data-Driven Documents, `http://www.d3js.org/`

[2] `http://www.astro.uni-bonn.de/~mtewes/d3cs/tdc1/` (See "Read me first" for help).

[3] This same category was named `doubtless` in Liao et al. (2015).

Table 1. `D3CS` classification of the TDC1 light curve pairs.

Rung 0	Estimates	ϵ
Secure	548 (53.5%)	0%
Plausible	291 (28.4%)	2.1%
Multimodal	60 (5.9%)	30.0%
Uninformative	125 (12.2%)	–
Rung 1	Estimates	ϵ
Secure	288 (28.1%)	0%
Plausible	383 (37.4%)	1.3%
Multimodal	127 (12.4%)	17.3%
Uninformative	226 (22.1%)	–
Rung 2	Estimates	ϵ
Secure	223 (21.8%)	0%
Plausible	406 (39.6%)	1.2%
Multimodal	168 (16.4%)	27.4%
Uninformative	227 (22.2%)	–
Rung 3	Estimates	ϵ
Secure	329 (32.1%)	0%
Plausible	324 (31.7%)	4.9%
Multimodal	161 (15.7%)	18.6%
Uninformative	210 (20.5%)	–
Rung 4	Estimates	ϵ
Secure	235 (23.0%)	0%
Plausible	430 (42.0%)	3.5%
Multimodal	108 (10.5%)	26.9%
Uninformative	251 (24.5%)	–
All Rungs	Estimates	ϵ
Secure	1623 (31.7%)	0%
Plausible	1834 (35.8%)	2.6%
Multimodal	624 (12.2%)	23.2%
Uninformative	1039 (20.3%)	–

Notes. The `D3CS` visual estimates for the time delays are shown for the 4 confidence categories defined in Sect. 3.1. The fraction of catastrophic outliers is given for each rung by ϵ, i.e., the time-delay estimations that are more than 20 days away from the truth.

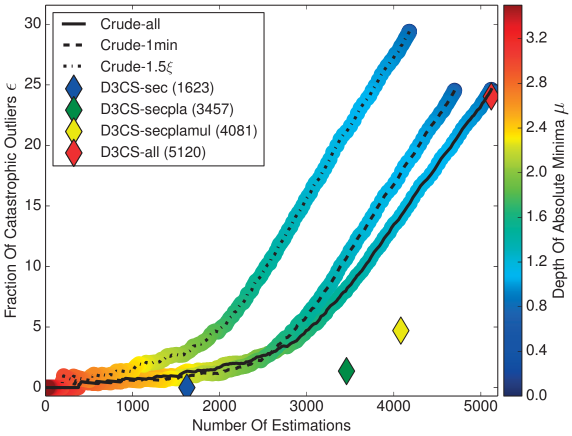


Fig. 2. Cumulative evolution of the fraction of catastrophic outliers ϵ (in percentage points) as a function of the number of time-delay estimations. To produce the plot, the curves are first sorted according to the depth of their absolute minimum μ, indicated in the colour bar. Each black line (solid, dashed and dotted) represents a different subsample (see text for details). The coloured diamonds show the value of ϵ for the `D3CS` combined samples; the corresponding number of estimations are indicated in parenthesis.

of TDC1. The concept of the method is to estimate a time delay by fitting a spline on one of the two light curves, and computing the residual signal of the second light curve relative to the spline after applying time and magnitude shifts. The details of the method are described in Appendix A; the present section evaluates its performance and compares this estimation to the visual time-delay values.

We characterize the efficiency of the automated Stage I procedure by comparing its fraction of catastrophic outliers ϵ with that of `D3CS`. We define catastrophic outliers as time-delay estimations deviating from the truth by more than 20 days, i.e. with $|\widetilde{\Delta t_i} - \Delta t_i| > 20$ days. The time-delay and confidence estimation evaluated by the automated procedure are reduced to two numbers: the depth of the absolute minimum μ and the interseason variations of the microlensing ξ. Figure 2 shows the evolution of the fraction of catastrophic outliers ϵ as a function of the cumulative number of time-delay estimations, sorted by increasing μ. The larger μ is, the more confident the automated procedure in the time-delay estimation is. We study the impact of the automated procedure parameters μ and ξ by introducing three subsamples of automated estimations:

- the `Crude-all` subsample contains all the estimations;
- the `Crude-1min` subsample contains only the estimations for which the procedure finds a unique local minimum with a depth μ <1;
- the `Crude-1.5`ξ subsample contains only the estimations with a magnitude shift deviation $\xi < 1.5$ at the location of the absolute minimum μ.

Figure 2 shows the fraction of outliers ϵ in the three subsamples and compares them to the value obtained visually with `D3CS`, which are shown as four diamond-shaped symbols corresponding to the combination of the four confidence categories of `D3CS` described in Sect. 3.1. We note that the uninformative `D3CS` estimations are systematically considered as catastrophic outliers here.

The selection criteria applied to the `Crude-1min` and `Crude-1.5`ξ subsamples are not able to decrease the fraction of outliers. This highlights how difficult it is to find efficient selection criteria for the automated procedure parameters, although no exhaustive exploration of the parameters space has been conducted. As expected, all the `D3CS` subsamples contain fewer outliers than the corresponding automated procedure subsamples. However, the efficiency of the latter are of the same order as the other automated methods presented in the TDC1 paper, which have $\epsilon = 2{-}3\%$ when half of the TDC1 data, i.e. 2500 light curve pairs, are measured.

In conclusion, although the automated procedure presented here is less complete and reliable than `D3CS`, it yields candidates that can be evaluated by eye in a second phase. Such a combined approach would benefit both from the speed of the automated method and from the flexibility of the human eye estimate when dealing with a broad variety of properties in the light curves. We note, however, that in the rest of the paper, we only use the results obtained via `D3CS` as our Stage I estimates.

4. Stage II: measuring time delays

Using the Stage I results as initial estimates, we proceed in this section by running our `PyCS` time-delay measurement techniques on the simulated TDC1 light curves. In Tewes et al. (2013a), three different algorithms were proposed: the simultaneous fit of the light curves using free-knot splines, the regression difference technique, and an approach based on a dispersion measurement of which the free-knot splines and the regression difference technique yielded the most accurate and precise results when applied to simulated COSMOGRAIL data (in Courbin et al. 2011; Tewes et al. 2013b; Eulaers et al. 2013; Rathna Kumar et al. 2013). To analyse the TDC1 simulations, we have therefore focused on adapting only these two most promising methods for an automated use.

We note again that our Stage II methods cannot be asked to judge the plausibility of a proposed delay. This step belongs to the Stage I method, i.e. to the visual inspection with `D3CS` to prevent or at least reduce catastrophic outliers. In practice, despite a correct Stage I estimate, any automated Stage II method may fail to converge, or it may yield a time-delay measurement that is incompatible with the initial approximation. To prevent these failures from contaminating our measurements we systematically discard any Stage II result that does not lie within 1.5 `D3CS` uncertainty estimate of the `D3CS` point estimate. This threshold acknowledges that the uncertainty estimates obtained from `D3CS` are typically overestimated by a factor of 2 to 3, which has been confirmed by Liao et al. (2015). We note that this rejection affects less than 1% of the light curve pairs and has no significant influence on the f metric.

4.1. Free-knot spline technique

In the free-knot spline technique (`spl`), each light curve in a pair is modelled as the sum of a spline representing the intrinsic variability of the quasar, common to both images of the pair, and an independent spline representing the extrinsic variability due to microlensing. The intrinsic spline has a higher density of knots and is therefore more flexible accomodating the quasar variability, which is assumed to be faster than the microlensing variability. During the iterative fitting process, the light curves are shifted in time so as to optimize the χ^2 between the data and the model light curve. To analyse a TDC1 light curve pair, we repeat this fit 20 times, starting form different initial conditions covering the Stage I uncertainty. This tests the robustness of the optimization procedure. The best model fit is then used to generate 40 simulated noisy light curves with a range of true time delays around the best-fit solution and using the temporal sampling of the original light curves. By blindly rerunning the spline fit on these simulated data, and comparing the resulting delays with the true input time delays, the delay measurement uncertainty is estimated.

We simplified and automated the `spl` algorithm for TDC1 with respect to the description of the free-knot spline method given in (Tewes et al. 2013a) and its applications to real COSMOGRAIL data. The main adaptations are the following:

1. The temporal density of spline knots controlling the flexibility of the intrinsic spline was computed from the signal-to-noise ratios measured on the two light curves, using an empirical calibration. The signal-to-noise ratios were obtained from a structure function, by comparing the typical amplitude of the light curve variability observed on a timescale of 50 to 75 days with the scatter between temporally adjacent observing epochs. For the microlensing spline, the knot density was fixed to be the same for all TDC1 pairs.
2. When generating our mock light curves, we did not inject any fast microlensing signal to mimic correlated noise. Only plain white noise was added to the generative model.
3. We did not analyse the time-delay measurement errors on the simulated curves as a function of true time delay. Instead, only the RMS error of these time-delay measurements was used as our total uncertainty estimate.
4. Finally, we did not manually fine-tune any parameters or correct for problematic model fits.

As a result, the entire `spl` analysis took about 5 CPU-minutes per TDC1 pair.

4.2. Regression difference with splines

Our second Stage II method, `sdi` (for spline difference), is based on the regression difference technique of Tewes et al. (2013a). To speed up the analysis, we replace the Gaussian process regressions by spline fits. In summary, the method independently fits a different spline to each of the two light curves, and then minimizes the variability of the difference between these two splines by shifting them in time with respect to each other. The advantage of this approach is that it does not require an explicit microlensing model. To estimate the uncertainty, the `sdi` method is run on the same simulated light curves provided by the `spl` technique. The `sdi` method has an even lower computational cost than `spl`.

5. Results on the TDC1 data

In this section, we separately analyse results from the Stage I and Stage II measurement processes as obtained on the simulated light curves of TDC1. General results from Liao et al. (2015) regarding submissions prepared with our methods include the following:

1. The Stage II methods `spl` and `sdi` show no significant deviations of the accuracy A from zero, and can thus be considered as inherently unbiased, given the statistical limits due to the finite challenge size.
2. The claimed precisions P of `spl` and `sdi` are very good, with χ^2 values of the order of $\chi^2_{\texttt{spl}} \simeq 0.5$ and $\chi^2_{\texttt{sdi}} \simeq 1.0$.
3. Based on results from the `spl` method simple power-law models for the dependence of A, P, and f on monitoring cadence, season length, and campaign length were adjusted. These relations attempt to capture general trends regarding the behaviour of all methods used in the context of TDC1, including our `spl` technique.

In the present paper, we focus on aspects that are complementary to the discussion of Liao et al. (2015).

5.1. Efficiency of time-delay discovery (Stage I)

We start by analysing the fraction of light curve pairs for which a time delay can be discovered with visual inspection, as a function of time delay and image brightness. This analysis relates only to the first stage of the time-delay measurement process.

Aside from the time delay and the quasar image brightness, the question of discoverability depends on observational conditions (e.g. monitoring cadence and duration) and on astrophysical characteristics (e.g. amount of quasar variability and

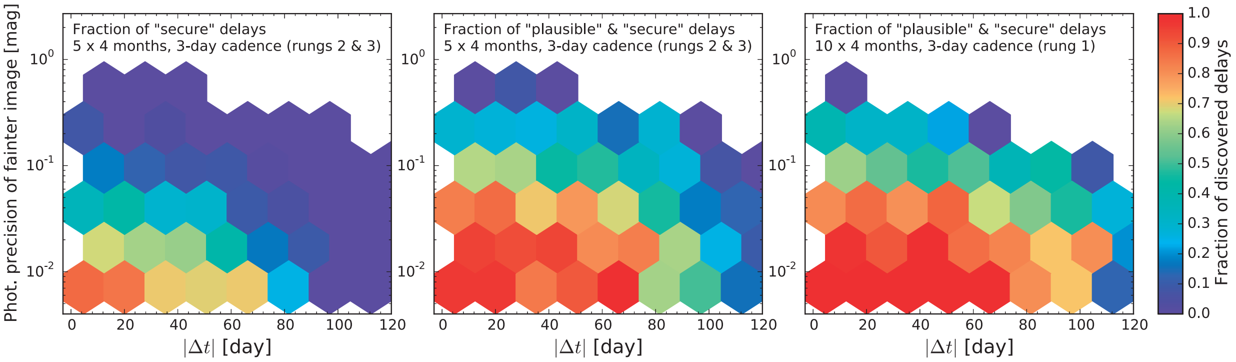


Fig. 3. Quantitative analysis of the discoverability of time delays through the extensive visual search with `D3CS` (Stage I) in the case of four-month observing seasons and a cadence of 3 days. The coloured tiles show the fraction of discovered delays as a function of the photometric precision of the fainter quasar image and the true time delay of the system. *Left panel* shows results from the very conservative `sec` sample, and *central panel* shows the less pure `secpla` sample that includes delay candidates considered as plausible (see text). *Right panel*, also for `secpla`, doubles the number of observing seasons. In each panel, only tiles covering more than three light curve pairs are shown.

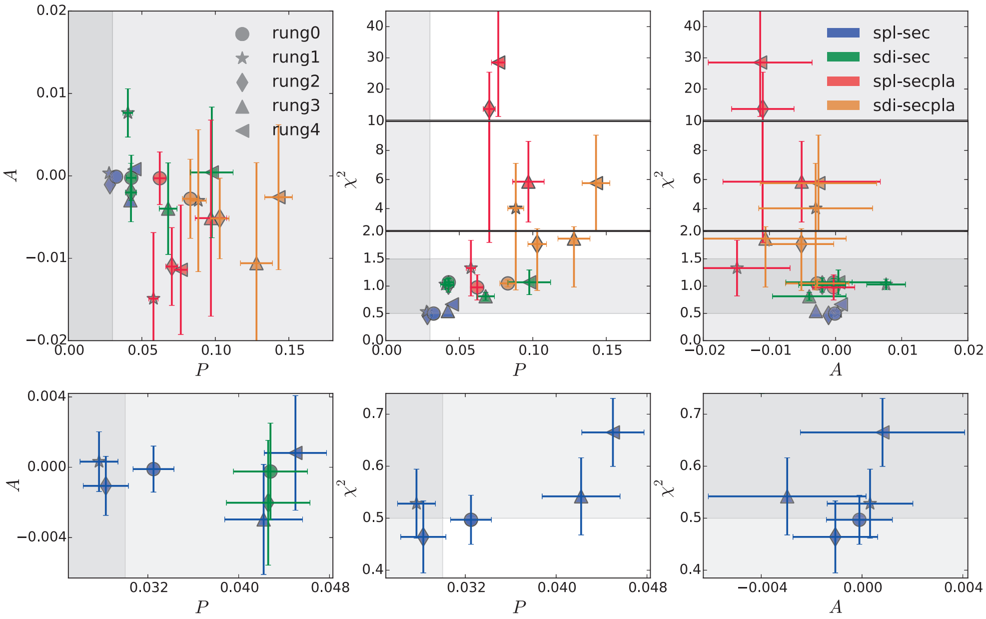


Fig. 4. Summary of metrics obtained with the Stage II algorithms `spl` and `sdi`, without any a posteriori clipping of the outliers. *Bottom row* presents enlargements taken from the panels on the *upper row*. The shaded regions represent the somewhat arbitrary target areas that were defined in the TDC1 paper.

microlensing perturbations). In the following, we select a given observing strategy and average over the astrophysical parameters of the TDC1 simulations, which follow clearly motivated distributions (Dobler et al. 2015). A large sample of light curve pairs with almost identical observing conditions can be obtained by merging rungs 2 and 3. These rungs share the fiducial three-day monitoring cadence for five seasons of four months each. The differing cadence dispersion of 0.0 days for rung 2 and 1.0 days for rung 3 (Table 1 of Liao et al. 2015) do not have a significant impact on the discoverability of time delays.

In practice, time delays can be measured accurately in pairs of light curves if the quality of both light curves is sufficient. In the following we consider as a relevant observable the photometric precision achieved in the fainter image of a pair. This is computed for each pair of light curves, as the median of the photometric error bars across the epochs of the TDC1 simulations. This is made legitimate by their overall effectiveness in representing the amplitude of the simulated noise, except for very few "evil" epochs of some systems (see Sect. 2.5 of Liao et al. 2015). However, when analysing real light curves, using the photometric scatter between the points might be a better choice than using potentially mis-estimated photometric error bars.

Figure 3 presents the distribution of the fraction of light curve pairs for which time delays could be identified via a meticulous `D3CS` visual inspection for two different monitoring strategies. In the left panel, only time delays categorized as secure through the visual inspection are considered as discovered. This is very conservative because in a real survey, simple lens models

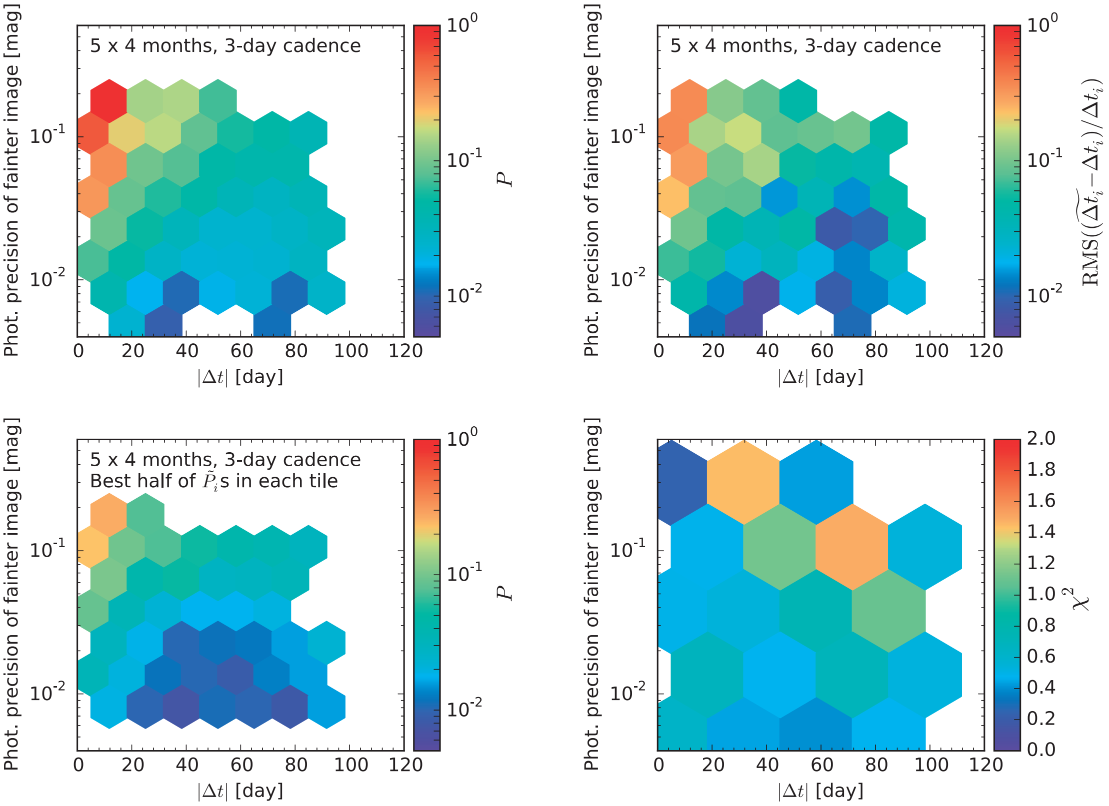


Fig. 5. Quantitative analysis of the precision achieved for the Stage II time-delay measurement as a function of the photometric precision of the fainter quasar image and as a function of the true time delay. All panels show results from the bias-free `spl` technique for the `secure` + `plausible` selection of rungs 2 and 3 after rejection of the catastrophic outliers (see text). *Top left panel* shows the metric P as computed using the uncertainty estimates δ_i without using $\widetilde{\Delta t_i}$. *Top right panel* shows the rms of the relative point estimation residuals without considering δ_i. *Bottom left panel* shows the average P obtained in each tile after selecting only the best half of systems according to the blind precision in $\tilde{P}$. *Bottom right panel* shows a map of the χ^2 metric. In all panels, only tiles describing more than three light curve pairs are shown.

will help to identify the correct time delay for almost all of the plausible systems as well. For this reason we also consider the combined `secpla` sample, shown in the central panel.

Some of the cases categorized as multimodal could certainly also be resolved using simple lens model considerations, but in practice the vast majority of these light curve pairs contain too few clear common features to estimate a reliable time delay, even if an approximate value would be known from the modelling. We therefore consider the discoverability of the `secpla` selection shown in the central panel of Fig. 3 as roughly representative of the fraction of potentially helpful delays that can be reliably measured from a real monitoring campaign or survey. It can be seen as an approximate lower limit for the fraction of time delays that can be observationally constrained in the absence of prior from a lens model, provided the properties of the microlensing signal are similar to those of the simulated light curves used here. Finally, the right panel shows the evolution of this discoverability if the same monitoring strategy is carried out for five more seasons, i.e., for a total of ten years.

We observe that after five years of monitoring, light curve pairs with a photometric precision in the fainter image better than $\sigma = 0.1$ mag and a true time delay shorter than $\Delta t = 80$ days (2/3 of the season length) are very likely to yield a time-delay measurement. Pursuing the monitoring for five more years significantly increases the average chances that longer delays up to ~90% of the season length become measurable.

5.2. Precision of time-delay measurement (Stage II)

We now turn to the analysis of the time-delay measurements (Stage II) for all systems where the time delay is successfully discovered (Stage I).

Figure 4 summarizes the results of the `spl` and `sdi` methods in terms of the metrics A (accuracy), P (claimed precision), and χ^2, as defined in Sect. 2. The figure follows the same conventions as Table 4 of Liao et al. (2015), but also includes measurements obtained on the `secpla` samples of each rung. As expected, the results for these `secpla` samples are more scattered than for the `sec` samples. The reason for these significant offsets in A and χ^2 with respect to the `sec` results is the stronger impact of outliers on the non-robust metrics.

To characterize the achievable precision of the Stage II measurements without being influenced by catastrophic outliers but still benefiting from a large number of light curve pairs, we now focus on the `secpla` sample from which we remove systems with estimated delays that are off by more than 20 days. This rejects about 1% of the `secpla` sample. We also note that catastrophic outliers are errors of the Stage I estimate, not Stage II.

Figure 5 presents metrics related to the Stage II time-delay measurement precision as a function of the photometric quality of the fainter quasar light curve and the time delay. In each tile the top left panel shows the average claimed precision P for the `spl` technique, for a five-year monitoring with four-month

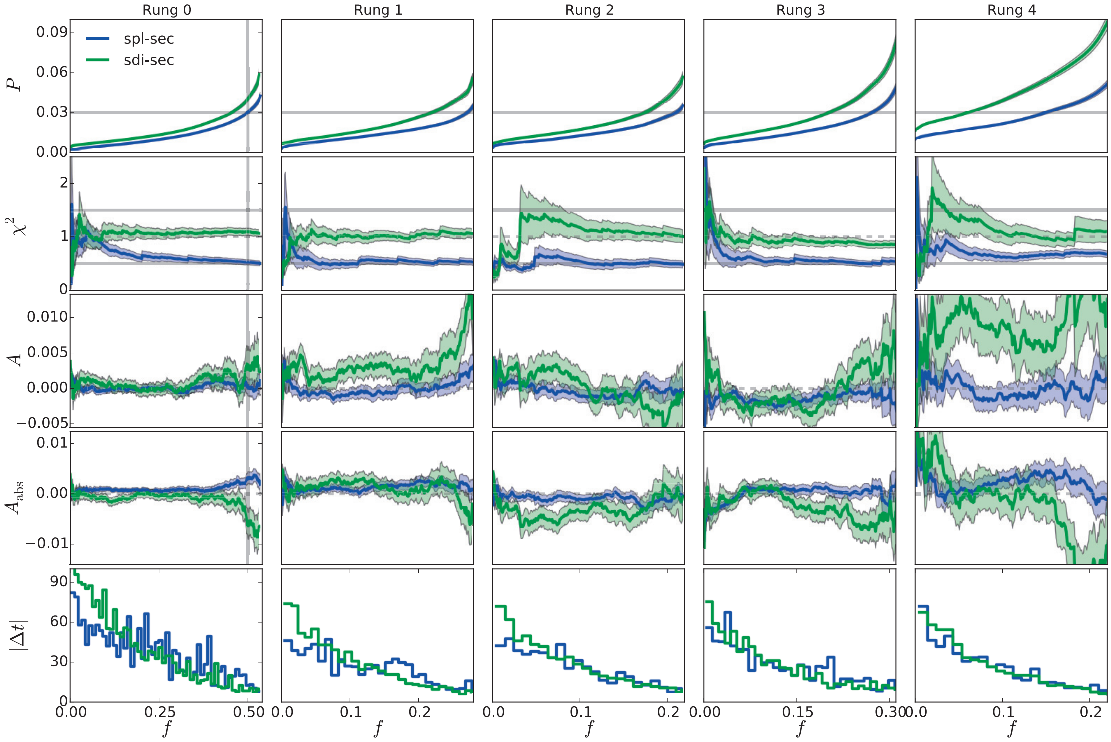


Fig. 6. Evolution of the TDC1 metrics per rung as a function of the fraction of estimations f for the `spl-sec` and `sdi-sec` samples. The plots are sorted by increasing values of the blind precision $\tilde{P}$ (see text). Shaded regions represent the error on the metrics. Solid grey lines show the target values for the metrics as defined in the TDC1 paper. Dashed grey lines show the best possible value for each metric. The *bottom row* presents the non-cumulative evolution of the median of the true time delays $|\Delta t_i|$ in bins of ten estimations.

seasons and a cadence of three days. We find that the cadence dispersion plays no significant role in this analysis, and we therefore merge rungs 2 and 3 to obtain this larger sample.

In contrast, each tile of the top right panel shows the root mean square (rms) of the relative error of the time-delay estimates $\widetilde{\Delta t}$. The observed structure is inevitably noisier because this rms is computed from the actual point estimates, while the precision P is based on the corresponding claimed uncertainty estimates. We observe both a qualitative and a quantitative similarity between these plots, suggesting that the time-delay uncertainty estimates, δ_i (Eq. (2)), from the `PyCS` techniques correctly capture trends related to the photometric quality and to the amount of temporal overlap in the light curves.

In the lower right panel of Fig. 5, the map of χ^2 metrics for the `spl` technique shows no strong evolution across the well-sampled regions of the parameter space. It does however indicate that the uncertainty estimates δ_i from `spl` are too conservative by a small factor of $(\chi^2)^{-1/2} \simeq 0.5^{-1/2} \simeq 1.4$. This is particularly visible for the high quality light curves with small time delays, i.e. with large temporal overlaps. Finally, the bottom left panel shows the average P metric computed using only the best half of light curve pairs in each tile, where the quality of a system is judged via the blind relative precision $\tilde{P}_i$ (see Eq. (5)). This operation, aimed at increasing the precision, divides by two the usable fraction of systems as given in Fig. 3. We consider such a selection in more detail in Sect. 5.3.

We observe in Fig. 5 that the best average relative precision in time-delay measurements seems to be achieved for time delays in the range 40−80 days for this particular monitoring strategy. This corresponds to about half of the season length, and results from the trade-off between absolute delay length and amount of overlap in the light curves.

Given the observed general aptitude of our time-delay uncertainty estimates, and thus P, to describe the actual point estimate errors committed by the `spl` technique, and the excellent competitiveness of `spl` compared to other time delay measurement techniques (see, e.g., Fig. 13 of Liao et al. 2015), we see the left panels of Fig. 5 as roughly representative of the precision that can be achieved by a state-of-the-art time-delay measurement method.

5.3. Effect on the overall metrics of selecting the best systems

In Fig. 6 we show the evolution of the overall average metrics as a function of the fraction of estimations f for the `spl-sec` and `sdi-sec` samples. The five columns represent the five rungs and the estimations are sorted according to their blind precision $\tilde{P}$, i.e the precision estimate from the TDC1 data prior to the unblinding of the true values. The non-cumulative median value of the true delays (bottom row) is computed on consecutive bins of ten estimations. The shaded regions represent the statistical uncertainties of the metrics $X_{\rm err}$ defined in Eq. (7). These uncertainties are too small in the top row to be distinguished from the curves.

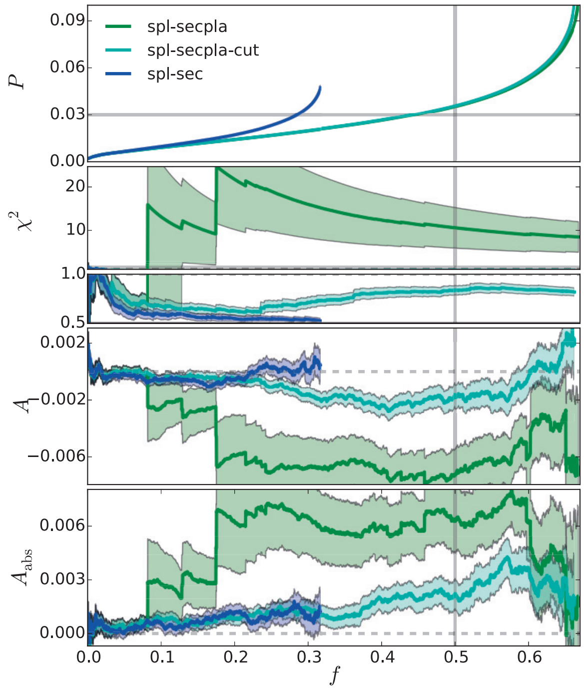


Fig. 7. Evolution of the TDC1 metrics with the fraction of estimations sorted by increasing blind precision $\tilde{P}$, for the `spl-sec` and `spl-secpla` samples merging all rungs. The `spl-secpla-cut` sample has been cleaned a posteriori from the outliers in the `spl-secpla` sample. In doing so, we removed 29 curves from the `spl-secpla` sample. The shaded regions, the solid and dashed grey lines are the same as in Fig. 6.

In the top row, P increases monotonically with f. This is expected since the estimations are sorted according to $\tilde{P}$ and since the `D3CS sec` sample is free of any outliers. The metrics χ^2, A, and $A_{\rm abs}$, respectively in the second, third and fourth rows stabilize around a mean value once a significant fraction of estimations is considered. The variations around the mean such as the jump in χ^2 at $f \sim 0.05$ in rung 2 are due to non-catastrophic outliers, i.e. time delays that deviate significantly from the truth but by less than 20 days. These outliers are the result of a non-optimal convergence of the Stage II methods for the curves with the lowest signal-to-noise.

The high-f end of A and $A_{\rm abs}$ are subject to strong variations in all rungs. These variations occur for small absolute values of the true time delay $|\Delta t|$. Similarly, the high-f end of P increases strongly. A small error on the time-delay estimation particularly affects P, A, and $A_{\rm abs}$ if the true time delay is small.

This correlation between the loss in precision and accuracy means that for the corresponding light curves, our algorithms estimate the time delays less accurately, but do provide larger error bars. We observe that the χ^2 remains constant as f increases. In conclusion, sorting the measurements in $\tilde{P}$ and rejecting a small fraction of the least precise estimations allows an optimal accuracy to be maintained without affecting the χ^2.

Figure 7 shows the evolution of the TDC1 metrics with the fraction of estimations, f, sorted by increasing order of $\tilde{P}$, for the `spl-sec` and `spl-secpla` sample. The few catastrophic outliers result in striking discontinuities in the curves. Quantifying the accuracy and precision of Stage II methods is different from avoiding such catastrophic outliers, and to address the former question, we also display in Fig. 7 a new subsample, `spl-secpla-cut`, where the 29 time-delay estimates with an absolute time-delay error larger than 20 days are a posteriori removed. Similarly, the impact of outliers can be reduced by considering the median of the individual metrics instead of their mean. This is not surprising, but nevertheless it reflects the need either to use metrics that can cope with outliers or, as in our Stage I approach, to make sure that no outliers contaminate the time-delay samples used for subsequent cosmological application.

6. Summary and conclusions

In this work, we used the simulated light curves of the TDC1 proposed by Dobler et al. (2015) to test the performance of the `PyCS` numerical techniques currently in use in COSMOGRAIL to measure time delays. These methods are fully data-driven, in the sense that they do not attempt to include any physical model for microlensing or for the intrinsic variations of quasars. This choice was deliberate and considers an empirical representation of the data that minimizes the risk of bias due to choosing the wrong physical model. The price to pay is that error bars on the measurements must be estimated from mocks and that we cannot use prior knowledge from external observations of quasars in a direct formal way. Using the same simulated light curves, we also assessed the quantity and quality of the time-delay measurements from future monitoring campaigns or surveys. We have made public our six main TDC1 submissions, obtained using the `D3CS`, `spl`, and `sdi` methods for the high purity `secure` and the less conservative `plausible` samples. These data are available on the COSMOGRAIL website[4]. Our results can be summarized as follows:

1. The visual estimation of time delays (Stage I) is extremely efficient in spotting catastrophic outliers and in providing useful time-delay estimates to be refined with subsequent numerical techniques (Stage II).
2. We attempted to build a simple automated time-delay estimation procedure that we can apply to the TDC1 data. While useful, this automated procedure does not achieve as good purity in the time-delay sample as the visual estimation. We note that we did not use this automated procedure for any of our submissions to the TDC1.
3. We provide a general analysis of the achievability of time-delay measurements as a function of the photometric precision of the light curves. In particular we show that almost all time delays shorter than two-thirds of the season length can be measured in five years of monitoring with four-month seasons and realistic photometric quality.
4. Our Stage II methods `spl` and `sdi` can be considered unbiased given the statistical limits due to the finite challenge size. The χ^2 values are close to unity. These good results emphasize the reliability of COSMOGRAIL time-delay measurements in general.
5. We quantify the average precision on the time-delay measurements as a function of photometric quality of the light curves. We find that the best average precision seems to be obtained for pairs whose time delay is approximately half of the monitoring season length.
6. We show that we can reliably evaluate the individual precision of our time-delay estimates. This may enable us, for any sample of light curves, to identify which are the most valuable objects to be followed up for cosmological purposes.

---

[4] `http://www.cosmograil.org/tdc1`

> We note, however, that any selection on the time delays in a quasar sample may also result in biases on the cosmological inference.

The above is true for the specific light curves of TDC1. These curves have been generated with simple models for quasar variations and microlensing and they do not include all potential nuisances of astrophysical, atmospheric, or instrumental origin. In addition, the PyCS techniques currently used in COSMOGRAIL do not attempt to account for these effects.

Acknowledgements. The authors would like to thank the organizers of the Time Delay Challenge, as well as Peter Schneider for useful discussions and the anonymous referee for his/her useful comments. This work is supported by the Swiss National Science Foundation (SNSF). Malte Tewes acknowledges support by a fellowship of the Alexander von Humboldt Foundation and the DFG grant Hi 1495/2-1. Dominique Sluse acknowledges support from a Back to Belgium grant from the Belgian Federal Science Policy (BELSPO).

References

Bostock, M., Ogievetsky, V., & Heer, J. 2011, IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)
Courbin, F., Chantry, V., Revaz, Y., et al. 2011, A&A, 536, A53
Dobler, G., Fassnacht, C. D., Treu, T., et al. 2015, ApJ, 799, 168
Eigenbrod, A., Courbin, F., Vuissoz, C., et al. 2005, A&A, 436, 25
Eulaers, E., Tewes, M., Magain, P., et al. 2013, A&A, 553, A121
Fadely, R., Keeton, C. R., Nakajima, R., & Bernstein, G. M. 2010, ApJ, 711, 246
Falco, E. E., Shapiro, I. I., Moustakas, L. A., & Davis, M. 1997, ApJ, 484, 70
Greene, Z. S., Suyu, S. H., Treu, T., et al. 2013, ApJ, 768, 39
Keeton, C. R., & Kochanek, C. S. 1997, ApJ, 487, 42
Liao, K., Treu, T., Marshall, P., et al. 2015, ApJ, 800, 11
Linder, E. V. 2011, Phys. Rev. D, 84, 123529
Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008, MNRAS, 389, 1179
Marshall, P. J., Lintott, C. J., & Fletcher, L. N. 2015, ARA&A, 53, 247
Molinari, N., Durand, J., & Sabatier, R. 2004, Comput. Stat. Data Analys., 45, 159
Moustakas, L. A., Abazajian, K., Benson, A., et al. 2009, ArXiv e-prints [arXiv:0902.3219]
Oguri, M., & Marshall, P. J. 2010, MNRAS, 405, 2579
Rathna Kumar, S., Tewes, M., Stalin, C. S., et al. 2013, A&A, 557, A44
Refsdal, S. 1964, MNRAS, 128, 307
Schneider, P., & Sluse, D. 2013, A&A, 559, A37
Schneider, P., & Sluse, D. 2014, A&A, 564, A103
Suyu, S. H., Treu, T., Blandford, R. D., et al. 2012, ArXiv e-prints [arXiv:1202.4459]
Suyu, S. H., Treu, T., Hilbert, S., et al. 2014, ApJ, 788, L35
Tewes, M., Courbin, F., & Meylan, G. 2013a, A&A, 553, A120
Tewes, M., Courbin, F., Meylan, G., et al. 2013b, A&A, 556, A22
Treu, T. 2010, ARA&A, 48, 87
Treu, T., & Koopmans, L. V. E. 2002, MNRAS, 337, L6
Weinberg, D. H., Mortonson, M. J., Eisenstein, D. J., et al. 2013, Phys. Rep., 530, 87

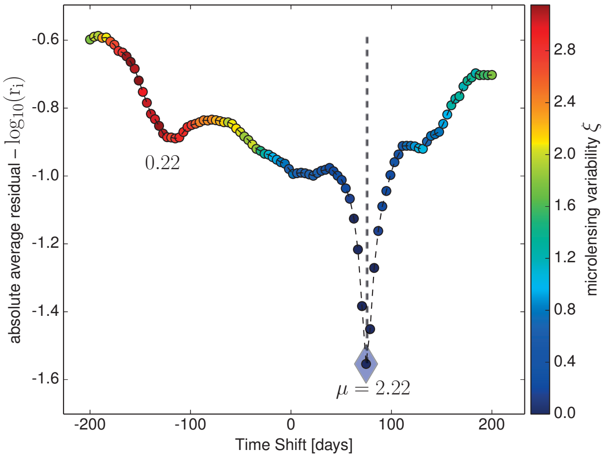


Fig. A.1. Example of time-delay estimations and confidence level from the automated procedure. The vertical grey dashed line represents the true value of the time delay. The blue diamonds correspond to the smallest absolute average residuals r_i returned by the automated procedure. The depth of the two minima μ_i are represented by the number below each minimum. The colour bar indicates the microlensing variability ξ (see text).

Appendix A: Automated Stage I procedure

If more time had been devoted to the visual inspection, we expect that more correctly estimated `plausible` time-delay estimations would have been classified as `secure`. After the TDC1, we developed an automated Stage I procedure. Its goal is to speed up and possibly improve the quality of the `D3CS` output by providing a range of reasonable initial time delays and associated confidence estimation. The following section describes technically how the time-delay and confidence estimation for this automated procedure are computed.

A.1. Time-delay estimation

For each pair of curves, a bounded-optimal-knot spline (Molinari et al. 2004; Tewes et al. 2013a) $s(t)$ is fitted to one of the two light curves. The second light curve $l(t)$ is then shifted by an amount δt in time and δm in magnitude. Thus, for a given observing epoch t, the value of the shifted light curve can be written as $l(t-\delta t)_{\delta \mathrm{m}}$. For every value of the time and magnitude shifts, we select all the N points in the second light curve that overlap in time with the spline. For these points, we compute the residuals res_n relative to the spline, i.e. the difference in magnitude between the points and the spline. The residual res_n for point n is

$$\mathrm{res}_n = [s(t) - l(t-\delta t)_{\delta m}]_n. \tag{A.1}$$

We then compute the average absolute residual $r(\delta t, \delta m)$ for every time and magnitude shift, i.e.

$$r(\delta t, \delta m) = \frac{1}{N} \sum_{n=1}^{N} \frac{|\mathrm{res_n}|}{\sqrt{N}}. \tag{A.2}$$

The possible presence of microlensing, assumed to be constant over an observing season, is handled in a very simple way. For each time shift δt_i, we apply independent magnitude shifts $\delta m_j(\delta t_i)$ to each season j. We define the residual curve $\boldsymbol{r} = \{r_1, ..., r_i, ..., r_\mathrm{T}\}$ as the sum of the smallest average absolute residuals for each season j. The i index runs from 1 to T and denotes the different time-delay values δt_i we want to test. This translates into

$$r_i = \sum_j \min_j \left[r(\delta t_i, \delta m_j) \right]. \tag{A.3}$$

We define r_i as a local minimum in $\boldsymbol{r}$ if

$$r_i < r_{i \pm k}, \quad \text{for } k = 1...10, \tag{A.4}$$

where we keep the absolute minimum in $\boldsymbol{r}$ as the final time-delay estimation. The k index running from 1 to 10 spans a range of ±20 days around each tested value r_i. Figure A.1 shows a typical residual curve $\boldsymbol{r}$ with the absolute minimum indicated as a coloured diamond and the true time delay indicated as a vertical dashed grey line.

A.2. Confidence estimation

For each pair of curves, we compute three parameters related to the shape of the residual curve $\boldsymbol{r}$ that can be used to estimate the quality of the time-delay estimations:

1. The number of local minima in $\boldsymbol{r}$.
2. The depth of each minimum μ_i and the absolute (i.e. the deepest) minimum μ,

   $$\mu_i = \frac{\bar{\boldsymbol{r}} - r_i}{\sigma_{\boldsymbol{r}}} \quad , \quad \mu = \min_{\delta t_i} [\mu_i], \tag{A.5}$$

   where σ_r is the standard deviation between the elements in $\boldsymbol{r}$. Examples of values for μ_i are indicated in Fig. A.1 below the minima for time shifts $\delta t \simeq -120$ days and $\delta t \simeq +80$ days.
3. The total magnitude shift $\boldsymbol{\delta m} = \{\delta m_1, ..., \delta m_\mathrm{T}\}$ and the microlensing variability $\xi(\delta t_i, \boldsymbol{\delta m})$, where we use the per season magnitude shifts $\delta m_j(\delta t_i)$ minimizing the average absolute residual r_i at each time shift δt_i,

   $$\delta m_i = \sum_j \delta m_j(\delta t_i) \quad , \quad \xi(\delta t_i, \boldsymbol{\delta m}) = \frac{\min\left[\boldsymbol{\delta m}\right] - \delta m_i}{\sigma_{\boldsymbol{\delta m}}}, \tag{A.6}$$

   where $\sigma_{\boldsymbol{\delta m}}$ is the standard deviation between the elements in $\boldsymbol{\delta m}$. In other words the quantity $\xi(\delta t_i, \boldsymbol{\delta m})$ is the difference between the sum of the magnitude shifts applied to each season at a given time shift δt_i and the minimum of this sum on all time shifts. This quantity follows the colour code in the sidebar of Fig. A.1 and is equivalent to the season-to-season microlensing variations minimizing the residuals for a given time shift. The lower this quantity is, the smaller the impact of microlensing is.

3.4 COSMOGRAIL time-delay measurements

Let us end this chapter about time-delay measurements by looking at the results of `PyCS` when applied on a selection of the COSMOGRAIL data. Only the time delays already published in the literature (or soon to be published) are presented here - other preliminary time-delay estimates for the other targets presented in Sec. 2.4 are saved for future publications.

3.4.1 HE0435-1223

Pay honor to whom honor is due. HE0435-1223 has been the target I worked on the most during the course of my PhD, and ended up playing a crucial role in the 3.8% precision measurement of H_0 (Bonvin et al., 2017) advertised in the title of this thesis. HE0435-1223 has already been used all along this chapter to illustrate the clockworks of `PyCS`. Indeed, Fig. 3.5 presents the intrinsic variance, Fig. 3.6 presents the time-delay error estimation and Fig. 3.7 synthesizes the previous figures by displaying the time delays and associated 1σ uncertainties, comparing the results of both the free-knot splines and regression difference techniques with the ones from Courbin et al. (2011).

What interests us here is to see how these results compare to other monitoring campaigns. Let us first have a look at the results from the pilot VLT monitoring campaign described in Sec. 2.5.1. Let us recall that the COSMOGRAIL cadence has been purposely augmented to one observation per night during ~1 month covering one of the gaps of the VLT monitoring campaign, in order to complement it. Figure 3.10 presents the results obtained when applying the free-knot splines technique on the FORS2 and ECAM data acquired between October 2014 and March 2015, and compare these results to the published measurements of Bonvin et al. (2017).

As expected, the results from FORS2 alone are not accurate at all, being often many days away from the Bonvin et al. (2017) results. On the opposite, the ECAM data acquired during the same period of time yields much more precise results, that are in excellent accordance with the results from the 13-years light curves. This highlights undeniably the crucial importance of a regular sampling during the monitoring campaign. Combining the two data sets results of estimates that are "in between" the results from the individual data set, consequently with larger error bars. Only on the AC and AD time delays does the combination of data sets yield more precise estimates than the ones from the individual data sets. This is not surprising, since it is only on these delays that both the FORS2 and ECAM estimates are roughly of the same precision. When an estimate is several times more precise than its counterpart from the other data set, the combination of the two data sets tend to average both the measured values and associated precision. There is nothing new in that, if not for the evidence that `PyCS` behaves as expected, which is always good to check. In the same regard, let us note that the less precise estimates nearly always encompass the more precise ones in their 1σ error bars, with the sole exception of the FORS2+ECAM AD estimate being at a 1σ tension with its Bonvin et al. (2017) counterpart. We interpret this concordance as a very positive sign of consistency

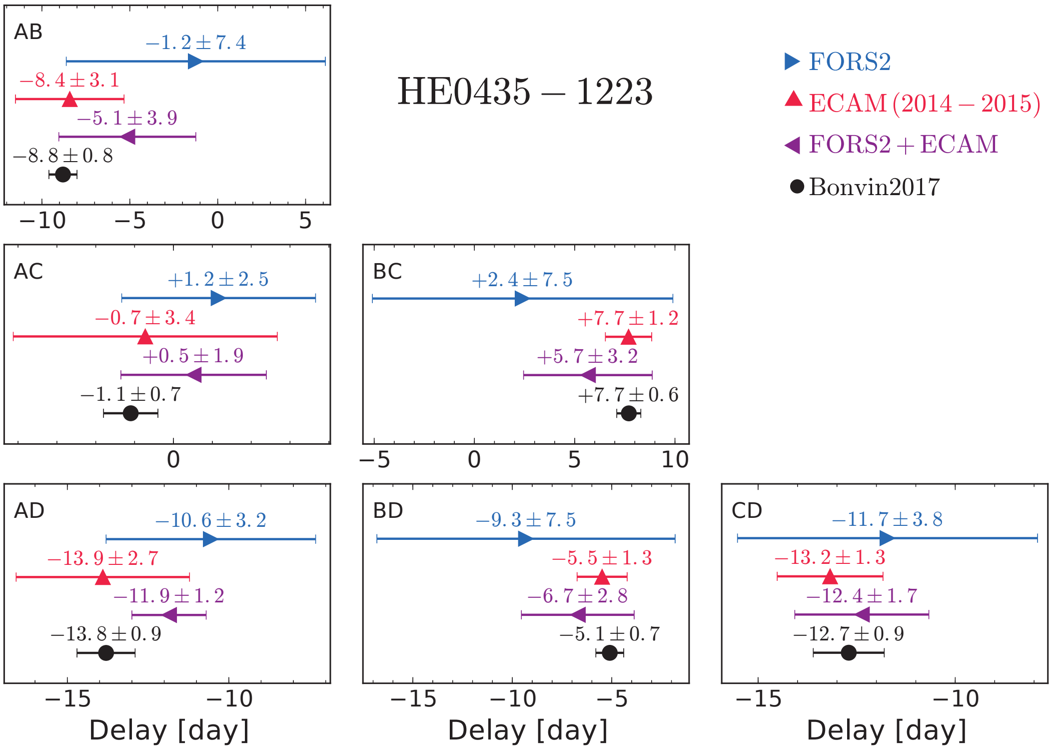


Figure 3.10: Time-delay measurements of the lensed quasar HE0435-1223 using six months of FORS2 data and ECAM data, acquired during the same interval of time. The colored results for FORS2, ECAM and FORS2+ECAM are obtained using the free-knot spline technique, with an initial knot step of $\eta = 20$ days for the splines modeling the intrinsic variations, and only three knots with fixed temporal position for the extrinsic variations. The results in black correspond to the free-knot splines technique applied to the 13-years COSMOGRAIL data presented in Bonvin et al. (2017).

in `PyCS`.

The COSMOGRAIL results are also marginally consistent with the results from Kochanek et al. (2006), that claimed a precision of less than a day on the AB, AC and AD delays - similarly to what is achieved here - but using only two observing seasons. The data used in Kochanek et al. (2006) and Bonvin et al. (2017) being approximately of similar sampling and quality during the 2004-2006 period, how can we then justify that it took ten more years to reach a similar precision in COSMOGRAIL? The answer lies in the confidence estimation procedure of `PyCS`, that is purposely certainly much more conservative than its existing counterparts (if any). It would make little sense now to try to apply the `PyCS` formalism to the Kochanek et al. (2006) data to see how it would have performed. It is more interesting, however, to study the evolution of the precision and apparent accuracy of the `PyCS` formalism with the addition of new monitoring seasons to see if our conservative approach is justified. By splitting the light curves of HE0435-1223 in three chunks (see the bottom part of each panel of Fig. 3.7), we see that the error bars are larger and consistent with the scatter of the time delays measured on the individual chunks. We interpret this as a strong signal towards the need of long monitoring campaign - at least at the COSMOGRAIL cadence and photometric precision - in order to avoid being over-precise or, in other words, under-accurate. Let us finally note that a recent reanalysis of the Courbin et al. (2011) data by Tsvetkova et al. (2016) that made use of polynomials and cross-correlations found time delays also marginally consistent with Bonvin et al. (2017).

To conclude, it is worth questioning if the regular COSMOGRAIL monitoring of HE0435-1223 should be extended beyond the 13-years landmark used in (Bonvin et al., 2017). In terms of cosmology, since Tie and Kochanek (2017) showed that short monitoring campaigns could be affected by systematic error in the time delays due to microlensing, going on for a few extra years would make no harm. In terms of time-delay measurements only, the answer is positive for the following reason: the analysis of HE0435-1223 for time-delay cosmography extended over several months, and time delays were estimated both without and with the most recent 2015-2016 monitoring season. The latter is particularly interesting, since it exhibits a strong variation in the intrinsic quasar luminosity (see Fig. 3.2). Such a season with so nicely apparent structures in the light curves improved the overall precision by $\sim 10\%$. In the current context where we aim at a $\sim 1\%$ precision on H_0, such an improvement is clearly beneficial; as a result, the monitoring of HE0435-1223 continues at the usual COSMOGRAIL four-days cadence. In addition, decade-long light curves give us through the modeling of microlensing events an unprecedented insight into the mechanisms at play. Fig. 3.2 hints towards currently ongoing microlensing events, possibly as important as the ones occurring in the 2007-2008 season. This speaks strongly in favor of continuing the monitoring.

3.4.2 HE0047-1756

HE0047-1756 has been monitored for more than 13 years by COSMOGRAIL, and for only 3 months at the 2m2. Although simulations predict that it is very likely that time-delay measurements acquired on such a short period of time are biased (see Sec. 2.5.6), we can nevertheless compare the time delays measured from Euler data (Fig. 2.11) and WFI data (Fig. 2.19). Note that a recent estimate of the time delay has been recently published by Giannini et al. (2017) from five years of monitoring data.

Let us start by applying the free-knot spline optimizer to the ECAM data. Since the quasar is faint, the light curves are sensibly more noisy than the ones of HE0435-1223, for instance. In order to no over-fit the data, we chose an initial knot step of $\eta_{\text{Euler}} = 60$ days to model the intrinsic variations. The resulting fit is presented on the top panel of Fig. 3.11. On the bottom panel of the same figure are the results of the free-knot spline techniques applied on the WFI data. The knot step for the intrinsic variations was $\eta_{\text{WFI}} = 15$ days, and no microlensing was considered.

It is interesting to see that the WFI data, once shifted, do not hint at all for the presence

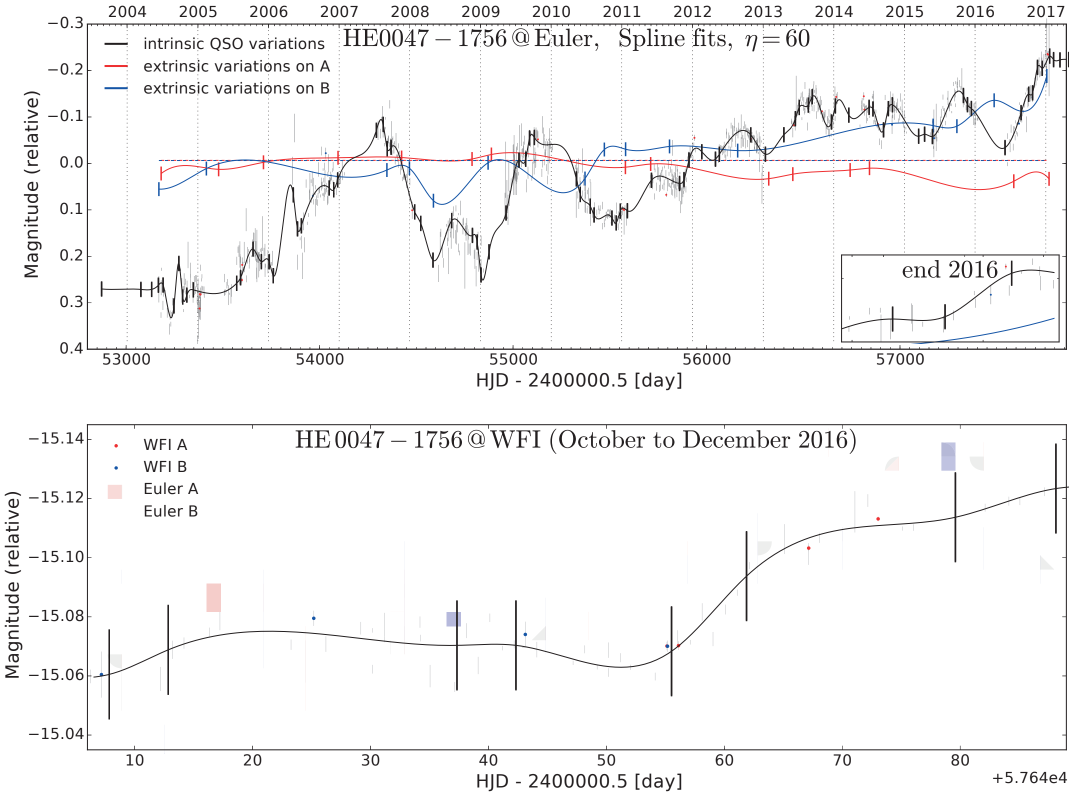


Figure 3.11: The top panel presents the results of the free-knot splines optimizer applied on 13 years of data acquired at the Euler telescope. Due to the relatively large photometric error bars, the knot step of the intrinsic spline has been chosen as $\eta_{\mathrm{Euler}} = 60$ days. The knot step of the extrinsic splines is $\eta_{\mathrm{Euler,ml}} = 300$ days. The bottom right insert zooms on the range MHJD=57646 to 57729, similar to the range displayed in the bottom panel. The bottom panel displays the results of the free-knot splines optimizer applied on ~3 months of WFI data. No microlensing was considered. The Euler data, correspondingly shifted in time and magnitude are over-plotted with larger symbols for comparison.

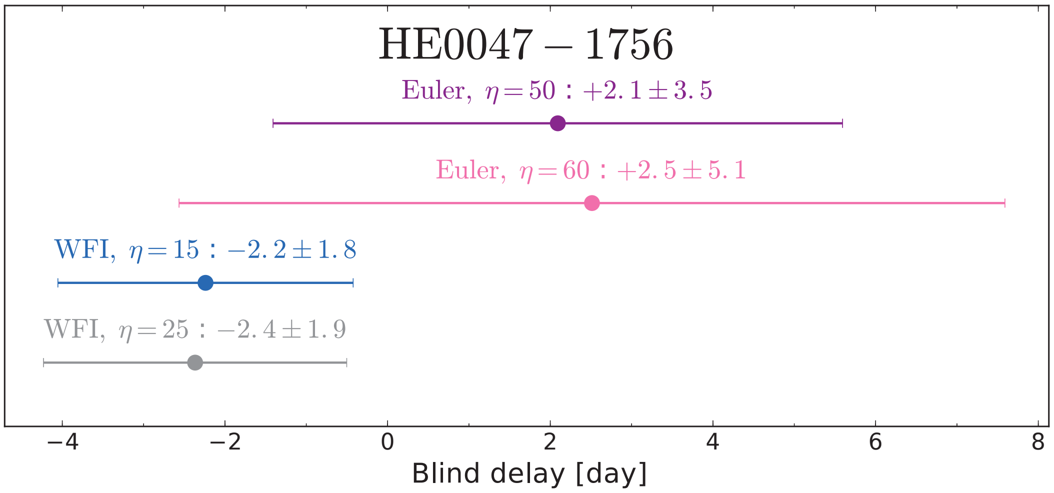


Figure 3.12: Blind time delays and associated 1σ error bars of HE047-1756 computed with the free-knot splines technique, independently on the Euler and WFI data sets. Microlensing was explicitly considered in the Euler data only, modeled through splines with an initial knot step of $\eta_{\text{Euler,ml}} = 300$ days. The so obtained time delays have been shifted by a constant (the mean of all estimates) to ensure blindness.

of microlensing. Coupled to the usual assumption we make that microlensing occurs only at large time scales, it becomes tempting to measure a time delay from the WFI data only, without explicitly considering any form microlensing. However, the ECAM data set speaks in favor of microlensing both in the A and B light curves during the same period of time. Figure 3.12 presents the blind time-delay measurements obtained by applying the free-knot splines technique on both data sets independently, with microlensing explicitly modeled in the Euler data but ignored in the WFI data. In each case, various knot steps for the intrinsic spline have been tested, which do not significantly affect the precision nor the accuracy of the results on the corresponding data sets.

The tension between the WFI and Euler time delays, although not very strong ($\sim 1\sigma$), anyhow points to an inconsistency somewhere in the analysis. Looking back to the original assumptions made prior to the application of the curve-shifting technique, the first thing that comes in mind is the absence of microlensing in the WFI component. Allowing for microlensing variability in the WFI data does not significantly change the time delay, yet increases its error bar by a factor of ~ 2. However, when forcing the microlensing modeled from the Euler data to be applied directly to the WFI data, the time-delay value from WFI shifts closer to the Euler value. We interpret this a a sign that although not directly visible in the WFI data, there is indeed microlensing that affects the time-delay measurement. The tricky thing is that assuming the microlensing modeled from the Euler data is correct, there is no way to reproduce it from the WFI data alone. In conclusion, if the current status definitely emphasizes the need for more monitoring data from WFI, it is also a striking real-life example of how easily we could have been biased with only the WFI data at hand. This converges towards the conclusion already established from numerical simulations (see Sec. 2.5.6) that a full monitoring season is needed.

3.4.3 DES J0408-5354

DES J0408-5354 is the first target monitored at the 2.2m ESO/MPG telescope to have the honors of an official publication. In only six month of monitoring with WFI, a time delay between the two brightest component has been measured with an estimated error of 2.1 days, corresponding to a precision of 1.8%. Figure 3.13 presents the best fit obtained by applying the free-knot splines optimizer on the light curves of the A, B and D images (C being too faint, it has been purposely excluded from the analysis), allowing for slow microlensing variability in these three components.

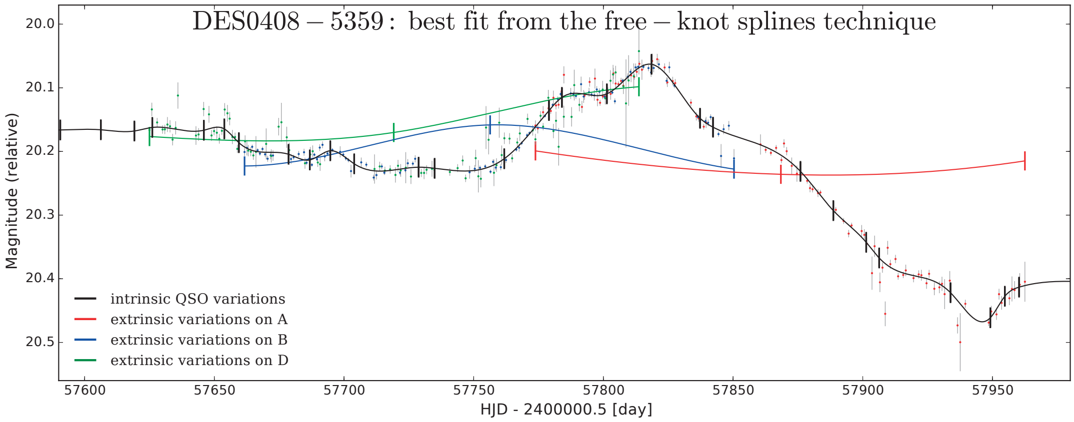


Figure 3.13: Shifted light curves of the lensed quasar DES J0408-5354 on which the free-knot spline optimizer with spline microlensing has been applied. Adapted from Courbin et al. (2017).

The very well defined peak around MHJD=57810 in Fig. 3.13 visible in the A and B light curves is the one allowing a very precise determination of the AB time delay. On the contrary, the AD and BD delays are very loosely constrained (if not constrained at all), with an estimated error of 29, respectively 32 days. These large uncertainties result from the faintness of image D, clearly not even approaching the mili-magnitude precision required by such short monitoring campaigns, which has two main consequences. First, the photometric error bars and the scatter around the best fit model of the D component are significantly larger than A and B. This results in the absence of fine and sharp structures in D that could be matched with equivalent structures in A or B. Second, D seems to be more significantly affected by microlensing; since the data display only one shallow inflection point (around MHJD=57750 on Fig. 3.13) occurring on a timescale similar to the modeled extrinsic variations (the green curve on Fig. 3.13), the contribution of D to the common intrinsic variations and its individual extrinsic variations are almost fully degenerate. Consequently, a large range of time delays involving the D light curve can be accommodated with a specific microlensing model.

In Courbin et al. (2017) - reproduced below -, we note that adding the ECAM data presented in Fig. 2.21 to the WFI data increases the uncertainty on the AB time delay, yet improve by a factor of ~ 2 the uncertainties of AD and BD time delays. This highlights well the importance of a high-cadence sampling and excellent photometric precision of the data; the well-marked peak in the A and B WFI light curves is much less constrained in the ECAM data, which in this case only add noisy measurements to the combination of the two data sets. However, a longer baseline of data points is also important. The monitoring of DES J0408-5354 at Euler started~ 3 months prior to the 2m2 monitoring, and these extra data, although not exhibiting any specific features, allow to significantly reduce the degeneracy between the modeled intrinsic and extrinsic variations in the D light curve.

In conclusion, the monitoring of DES J0408-5354 reflects well both the gap that has been crossed by the quasi nightly monitoring and the challenge it represents: being sensitive to the short and small-amplitude intrinsic luminosity variations of the quasar that are not degenerated with the extrinsic microlensing variability, like the peak in AB, allows to precisely measure time delays in only one season. Yet, failing to catch such short variations immediately results in extremely large uncertainties.

Astronomy & Astrophysics manuscript no. paper˙0408˙final
June 28, 2017

DES-2017-0252
Fermilab PUB-17-204-AE

COSMOGRAIL XVI: Time delays for the quadruply imaged quasar DES J0408−5354 with high-cadence photometric monitoring

F. Courbin[1], V. Bonvin[1], E. Buckley-Geer[2], C.D. Fassnacht[3], J. Frieman[2,4], H. Lin[2], P.J. Marshall[24], S.H. Suyu[6,7,8], T. Treu[9], T. Anguita[10,11], V. Motta[12], G. Meylan[1], E. Paic[1], M. Tewes[13], A. Agnello[14], D.C.-Y. Chao[6], M. Chijani[10], D. Gilman[9], K. Rojas[12], P. Williams[9], A. Hempel[10], S. Kim[15,16], R. Lachaume[15,16], M. Rabus[15,16], T. M. C. Abbott[17], S. Allam[2], J. Annis[2], M. Banerji[18,19], K. Bechtol[20], A. Benoit-Lévy[21,22,23], D. Brooks[22], D. L. Burke[24,25], A. Carnero Rosell[26,27], M. Carrasco Kind[28,29], J. Carretero[30], C. B. D'Andrea[31], L. N. da Costa[26,27], C. Davis[24], D. L. DePoy[32], S. Desai[33], B. Flaugher[2], P. Fosalba[34], J. García-Bellido[35], E. Gaztanaga[34], D. A. Goldstein[36,37], D. Gruen[24,25], R. A. Gruendl[28,29], J. Gschwend[26,27], G. Gutierrez[2], K. Honscheid[38,39], D. J. James[40,17], K. Kuehn[41], S. Kuhlmann[42], N. Kuropatkin[2], O. Lahav[22], M. Lima[43,26], M. A. G. Maia[26,27], M. March[31], J. L. Marshall[32], R. G. McMahon[18,19], F. Menanteau[28,29], R. Miquel[44,30], B. Nord[2], A. A. Plazas[45], E. Sanchez[46], V. Scarpine[2], R. Schindler[25], M. Schubnell[47], I. Sevilla-Noarbe[46], M. Smith[48], M. Soares-Santos[2], F. Sobreira[49,26], E. Suchyta[50], G. Tarle[47], D. L. Tucker[2], A. R. Walker[17], and W. Wester[2]

(Affiliations can be found after the references)

June 28, 2017

ABSTRACT

We present time-delay measurements for the new quadruply imaged quasar DES J0408−5354, the first quadruply imaged quasar found in the Dark Energy Survey (DES). Our result is made possible by implementing a new observational strategy using almost daily observations with the MPIA 2.2m telescope at La Silla observatory and deep exposures reaching a signal-to-noise ratio of about 1000 per quasar image. This data quality allows us to catch small photometric variations (a few mmag rms) of the quasar, acting on temporal scales much shorter than microlensing, hence making the time delay measurement very robust against microlensing. In only 7 months we measure very accurately one of the time delays in DES J0408−5354: $\Delta t(AB) = -112.1 \pm 2.1$ days (1.8%) using only the MPIA 2.2m data. In combination with data taken with the 1.2m Euler Swiss telescope, we also measure two delays involving the D component of the system $\Delta t(AD) = -155.5 \pm 12.8$ days (8.2%) and $\Delta t(BD) = -42.4 \pm 17.6$ days (41%), where all the error bars include systematics. Turning these time delays into cosmological constraints will require deep HST imaging or ground-based Adaptive Optics (AO), and information on the velocity field of the lensing galaxy.

Key words. methods: data analysis – gravitational lensing: strong – cosmological parameters

1. Introduction

Accurate and precise measurements of the time delay(s) between the multiple images of gravitationally lensed quasars offer an independent way of constraining cosmology. The method is simple and is mostly sensitive to H_0 with weak dependence on other cosmological parameters (Refsdal 1964). For this reason, the time-delay method has the potential to alleviate the degeneracies between cosmological parameters other than H_0. In addition, it provides helpful input to resolve the tension between H_0 as measured by Planck assuming a flat ΛCDM model (Planck Collaboration et al. 2016) and the local distance ladder, i.e. Cepheid stars (Freedman et al. 2001) and type Ia supernovae (e.g. Riess et al. 2016). Currently the results from Cepheids and supernovae are in tension with the CMB given the error bars (e.g. Freedman 2017). Quasar time delays offer an opportunity to measure H_0 completely independentely of any of the above probes.

The method requires several ingredients: i) time-delay measurements, ii) models constraining the mass and light distribution in the lensing galaxy, iii) an estimate of the contribution of objects along the line of sight to the overall potential well. The first point has been addressed by the COSMOGRAIL program, started in 2004 and delivering since then some of the best-quality time-delay measurements (e.g. Bonvin et al. 2017; Rathna Kumar et al. 2013; Tewes et al. 2013b; Courbin et al. 2011; Vuissoz et al. 2008; Courbin et al. 2005; Eigenbrod et al. 2005). In parallel, detailed modeling techniques have been developed and used on deep Hubble Space Telescope (HST) images in combination with spectroscopic data providing crucial constraints on the dynamics of the lensing galaxy (Treu & Koopmans 2002; Suyu et al. 2006, 2009). Such models, in combination with an estimate of the overall mass along the line-of-sight (e.g. Hilbert et al. 2009; McCully et al. 2017, 2014; Collett et al. 2013) allow one to measure the time-delay distance and consequently the Hubble parameter, H_0 (e.g. Suyu et al. 2010).

In order to perform precise cosmological measurement with strongly lensed quasars, these three ingredients must be accurately constrained. This has become possible only recently, with the joint efforts of the COSMOGRAIL (e.g. Courbin et al. 2005) and H0LiCOW programs (H_0 Lenses in COSMOGRAIL's Wellspring; Suyu et al. 2017), focusing on 5 well selected bright lensed quasars. Recent results can be found in Bonvin et al. (2017); Wong et al. (2017); Rusu et al. (2017); Sluse et al. (2016) who infer $H_0 = 71.9^{+2.4}_{-3.0}$ km s^{-1}Mpc^{-1} from 3 of the H0LiCOW lenses in a flat ΛCDM universe.

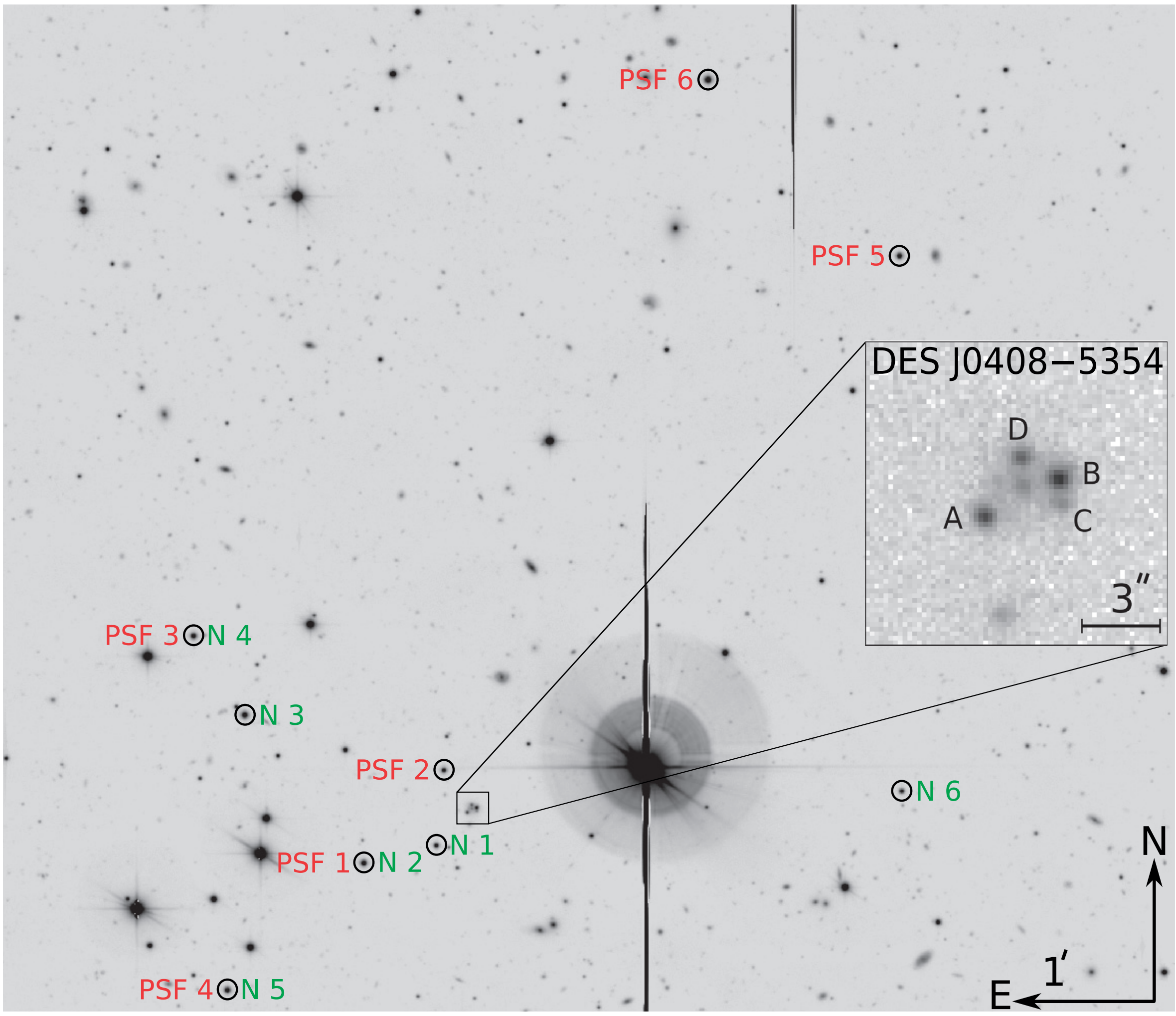


Fig. 1. Part of the field of view around DES J0408−5354 as seen with the 2.2m MPIA/ESO telescope at La Silla observatory. The image is a stack of 150 frames totaling 25h of exposure. The 6 PSF stars used to obtain the photometric light curves with deconvolution photometry are indicated in red. The 6 stars used for the frame-to-frame calibration of the relative photometry are indicated in green. The inset shows a 10″ zoom on DES J0408−5354 and is extracted from a single 640 sec exposure with 0.6″ seeing with the same labeling for the quasar images as in Fig. 1 of Lin et al. (2017). Note that image C is blended with a foreground lensing galaxy labelled G2 in Agnello et al. (2017) and Lin et al. (2017).

The H0LiCOW sample currently under study includes 5 lenses, with an expected H_0 measurement to <3.5% including systematics (Suyu et al. 2017). Going beyond this will require mass production of time delays. With 55 new time delays and dynamical measurements for the lensing galaxy Jee et al. (2015) estimate that H_0 can be measured to close to 1%. An independent study by Shajib, Agnello & Treu (2017, in prep) show that with resolved kinematics of the lens (e.g. with JWST or ground-based AO) 1% accuracy on H_0 can be reached with 40 lenses. This requires the discovery of new lenses, which is underway in the Dark Energy Survey (DES) (Ostrovski et al. 2017; Lin et al. 2017; Agnello et al. 2015), deep spectroscopy, the characterisation of the line-of-sight matter distribution, and the measurement of the time delays to a few percents for each individual system. The latter is the goal of the present work.

Because the slow intrinsic variations of the quasar occur roughly at the same time scale as the extrinsic variations (i.e. microlensing), measuring time delays requires years of monitoring. As the future of time delay cosmography resides in the measurement of several tens of new time delays, each time delay must be measured quickly, much faster than the typical 10 years needed with current lens monitoring campaigns. Current lens monitoring campaigns, including COSMOGRAIL, use 1m-class telescopes with a monitoring cadence of about 1 epoch per 3-4 days. The typical photometric accuracy with such data is limited to about 0.01 mag rms for many targets, hence allowing to catch only the most prominent features of the quasar variations. It is difficult or impossible to sufficiently disentangle these from extrinsic variations related to microlensing unless very long light curves are available (e.g. Bonvin et al. 2016; Liao et al. 2015).

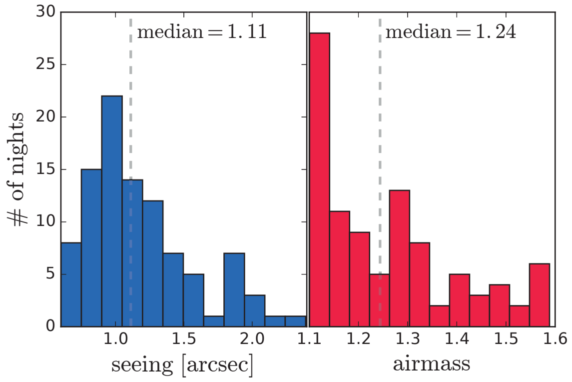


Fig. 2. Seeing and airmass distributions for the 7 months of observations of DES J0408−5354 with the WFI instrument on the MPIA 2.2m telescope.

In the present work, we implement a new high-cadence and high-signal-to-noise (high-S/N) lens monitoring program, with the goal of measuring time delays in only 1 single observing season rather than about 10 years as previously. With on average one observing point per day and a S/N of the order of 1000 per quasar component, we can now catch much faster variability in the intrinsic light curve of the quasar (e.g. Mosquera & Kochanek 2011). In almost all cases, these features on time scales of a few days to a few weeks are more than an order of magnitude faster than the extrinsic variations. This difference in signal frequencies makes it possible to disentangle much better between extrinsic and intrinsic quasar variations. As the small and fast quasar variations are frequent (e.g. Mushotzky et al. 2011), only a short monitoring period is required to measure time delays. The required temporal sampling and S/N are achieved by using the MPIA 2.2m telescope and the Wide Field Imager (WFI) at ESO La Silla Observatory daily, through a dedicated long-term monitoring program.

We present here our first time delay measurement obtained with the MPIA 2.2m telescope, for the quadruply-imaged quasar DES J0408−5354, at $z_q = 2.375$. DES J0408−5354 was identified as a quadruply imaged quasar by Lin et al. (2017). The lensing galaxy has a redshift of $z_l = 0.597$, measured by (Lin et al. 2017) using the Gemini-South telescope. Agnello et al. (2017) provide simple models for DES J0408−5354 using a deep image of the lens obtained from WFI data and predict time delays for a ΛCDM cosmology and different mass distributions including potential companions to the lensing galaxy, which influence the time-delay predictions.

2. Observations and photometry

The observational material for the present time-delay measurement consists of almost daily imaging data with the MPIA 2.2m telescope and of bi-weekly imaging with the 1.2m Euler Swiss telescope, both at ESO La Silla.

We started the observations on 1 October 2016 with the MPIA 2.2m telescope at ESO La Silla to monitor DES J0408−5354 through the R_c filter. The WFI instrument, mounted in the 2.2m telescope, has a total field of view of 36′ × 36′, covered by 8 CCDs with a pixel size of 0.2″. For our monitoring purpose we use only 1 chip to ensure a stable night-to-night calibration. This chip has a field of view of 9′ × 18′. Part of this it is shown in Fig. 1.

The WFI was used almost daily until 8 April 2017, i.e. over a total of 7 months of visibility of the object, except for 14 consecutive nights between 10 December 2016 and 24 December 2016 due to technical problems and for 1 week in January 2017 due to an extended period of poor weather. For each observing epoch, 4 dithered exposures of 640 sec each were taken in the R_c filter. A total of 459 images were taken in 7 months, of which we use 398 which have adequate seeing and PSF quality. On average, the resulting temporal sampling was one observing point every 1.96 day. The median seeing over this period was 1.1″. Thanks to the flexible scheduling of the observations at the telescope it was possible to observe DES J0408−5354 most of the time at low airmass. The seeing and airmass distributions of the observations are given in Fig. 2.

The high-cadence and high-S/N (2-3 mmag rms per quasar image) obtained with the 2.2m telescope allow us to catch much smaller and much shorter photometric variations than the COSMOGRAIL observations obtained with smaller 1m-size telescopes. We can typically see signals as small as a few mmags and as short as a 15-20 days, which is crucial to avoid contamination by extrinsic variations, as illustrated in Sect. 3

The data from the 1.2m Euler telecope were obtained in the R-band with the ECAM instrument from July 2016 to April 2017. The pixel size of the camera is 0.238″, providing a field of view of 14′ on-a-side. We took, for each of the 45 observing epochs 6 exposures of 360 sec each, i.e. 36 min in total. The mean temporal sampling for the Euler observations is of only 1 point every 5 days, but the Euler observations started about 100 days before the WFI observations, hence extending the length of the light curves.

The data reduction procedure applied to the images follows the standard COSMOGRAIL pipeline, as applied to the data obtained with the 1.2m Euler telescope for RX J1131−123 (Tewes et al. 2013b) and HE 0435−1223 (Bonvin et al. 2017). It includes subtraction of a bias level and flat-fielding using sky flats taken on average every few nights. Each frame is then sky-subtracted using the `GLOBAL` option in the `Sextractor` package (Bertin & Arnouts 1996). The data from the 2.2m telescope have significant fringe patterns on bright nights. We therefore construct a fringe image by iteratively sigma-clipping the four dithered exposure of each night and by taking the median. This image is then subtracted from the individual dithered exposures taken each night which are subsequently registered to the same pixel grid.

We carry out the photometric measurements using the deconvolution photometry with the so-called MCS image deconvolution algorithm (Magain et al. 1998; Cantale et al. 2016). This algorithm first computes a deconvolution kernel from the images of stars. The kernel is chosen so that the Point Spread Function (PSF) in the deconvolved images is a circular Gaussian function with a Full-Width-Half-Maximum (FWHM) of 2 pixels. The pixel size in the deconvolved images is half that of the original data, i.e. the resolution in the resulting images is 0.2″ for WFI and the pixel size is 0.1″. We show in Fig. 1 the PSF stars used as well as the reference stars for the image-to-image flux calibration.

The MCS algorithm deconvolves all the registered images simultaneously, i.e. each one with its own PSF. However all images share the same deconvolved model, which is decomposed into a point-source channel (quasar images) and an extended-channel (lensing galaxy and faint quasar host galaxy). In this process, the position of the point sources is the same for all im-

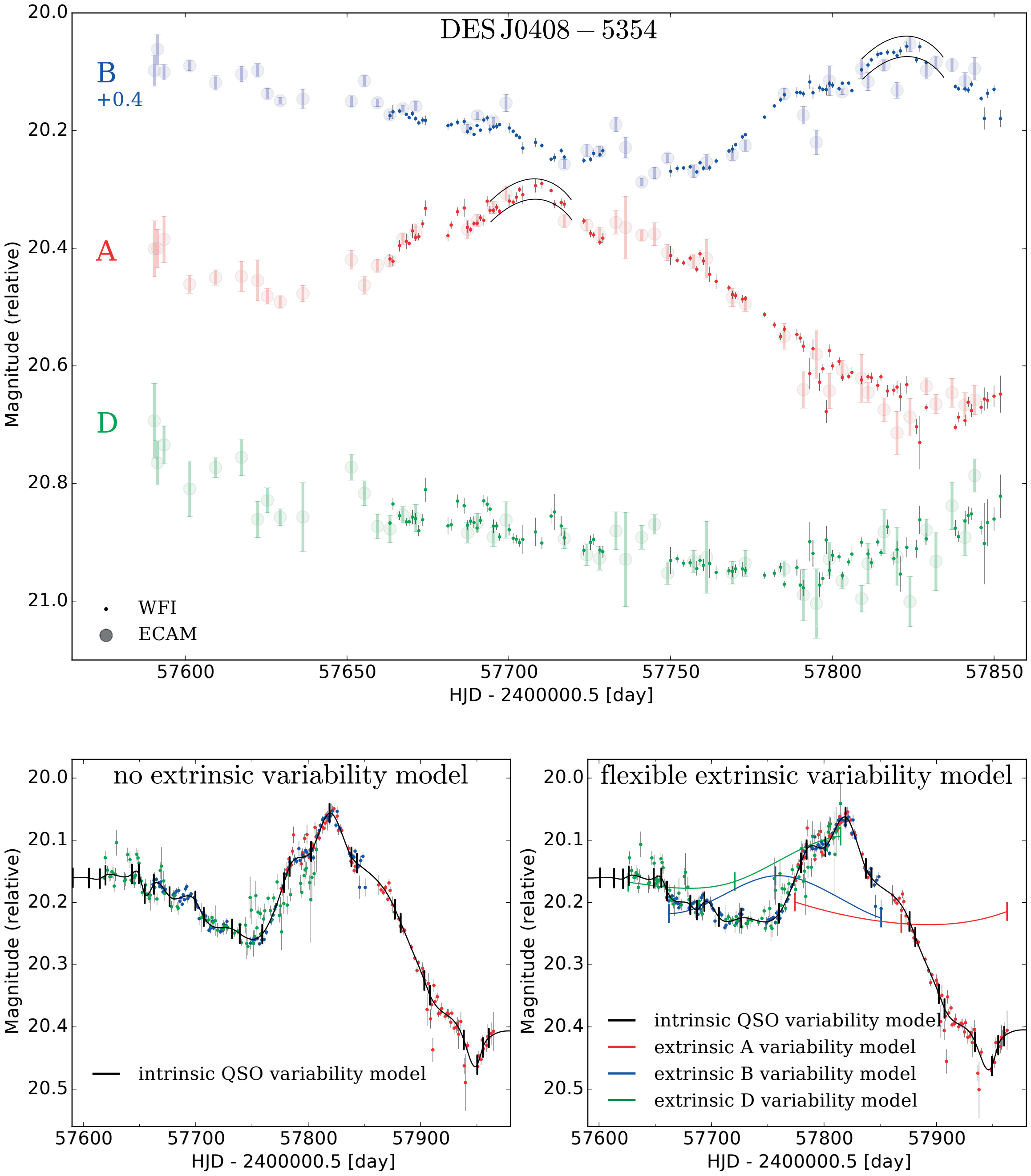


Fig. 3. *Top:* light curves for DES J0408−5354 obtained in the R_c filter with the MPIA 2.2m telescope and the WFI instrument. The points obtained with the 1.2m Euler telescope are also shown, with larger and thicker symbols. To guide the eye, the structure constraining the most the AB time delay is indicated between black solid lines. *Bottom left:* spline-fitting of the intrinsic quasar variations (with an initial knot step of η = 15 days) and time delay determination when neglecting extrinsic variation due to microlensing. Note that the time delay values do not depend much on the chice of this η parameter which is only an initial value optimized during the fit. *Bottom right:* same as bottom left but now including extrinsic variations (color curves). For more clarity the lower S/N Euler data are shown only in the upper left panel. Our light curves are available online at CDS and on `cosmograil.org`.

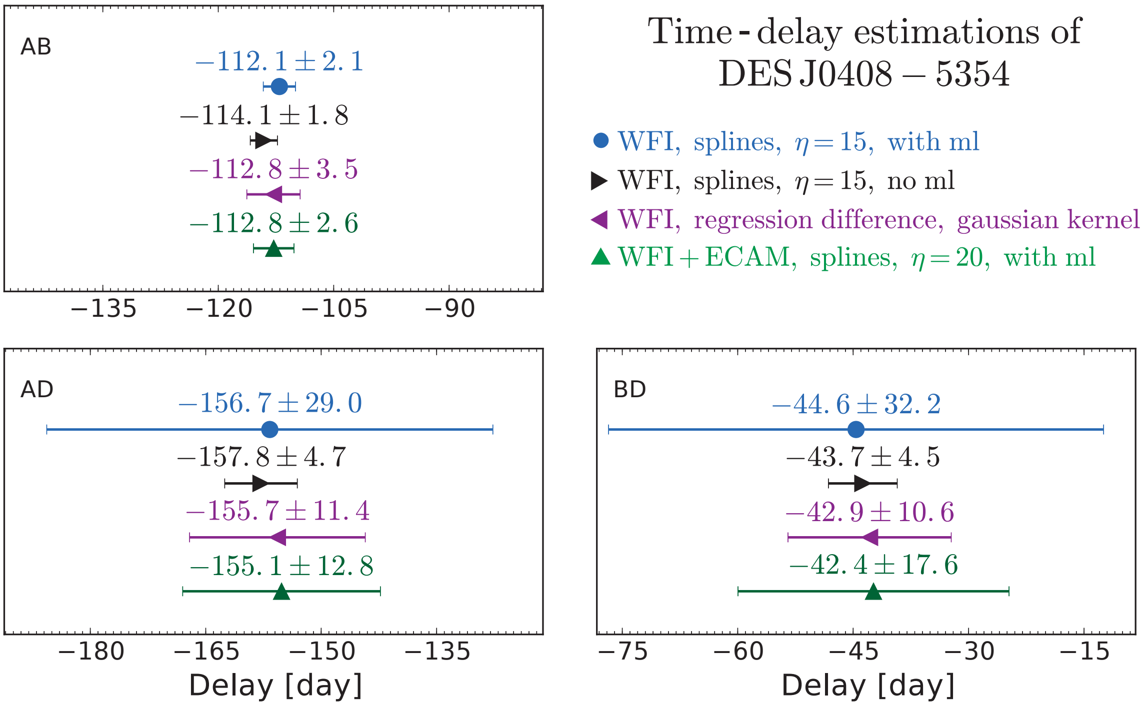


Fig. 4. Time delay measurements for the 3 brightest quasar images of DES J0408−5354 using the data shown in Fig. 3. The time delay measurements are carried out in 3 different ways. In blue are shown the results using only the WFI data and the spline fitting method and in purple with the regdiff method. In green are shown the results for the combined WFI+Euler data set using the spline fitting method. For comparison we also show, in black, the spline fitting result when using only the high-cadence WFI data and no model for the microlensing extrinsic variations. A negative $\Delta t(AB)$ means that signal from image A reaches the observer's plane before B.

ages as well as the extended-channel but the intensities of the point sources vary from image to image, hence leading to the photometric light curves. The latter are presented for the three brightest lensed images of DES J0408−5354 in Fig. 3 which shows the striking difference in depth and sampling between the 2.2m data and the Euler ones. Yet, the two datasets agree and complement each other well and exhibit fine structures in the light curves of A and B. The D component, however, has much shallower variations given the larger photometric error bars and will need further monitoring.

3. Time delay measurement

We use `PyCS`[1], a toolbox containing several algorithms to measure time delays from quasar light curves and accounting for the intrinsic variations (from the lensed source) and extrinsic variations (microlensing) in the data. The (public) algorithm was proposed by Tewes et al. (2013a) and tested on simulated data from the Time Delay Challenge (Bonvin et al. 2016; Liao et al. 2015), with overall excellent performances.

3.1. Time delay measurement with `PyCS`

`PyCS` is the stardard curve-shifting toolbox of the COSMOGRAIL project. We apply it to the WFI light curves as well as on the combination of the Euler and WFI data. We do not attempt to measure a time delay using the Euler data on their own as they contain only 45 epochs over the duration of the observations. However, used in combination with the WFI images, they increase the time baseline with observations between July 2016 and October 2016.

We use the two best algorithms of the `PyCS` toolbox: the *free-knot spline* technique and the *regression difference* technique (Tewes et al. 2013a). In the former, the intrinsic and extrinsic variations in the light curves are modelled explicitly as spline functions. In doing so, we give more flexibility, through the η parameter (i.e. initial knot spacing; Tewes et al. 2013a), to the spline representing the intrinsic variations of the quasar than to the spline representing extrinsic variations. The residuals to the fit when using extrinsic variations indicate that modeling microlensing with low flexibility for the knot positions is sufficient given the data. In practice, this is done by imposing that the spline representing the extrinsic variations has only 3 knots and that the central knot stays centered between the light curve extrema during the minimization process. In the *regression difference* technique, we minimize the variability of the difference between Gaussian-process regressions performed on each light curve individually. This second method has no explicitly parametrized form for the extrinsic variability, which makes the two techniques fundamentally different and independent. We apply the two methods to our data in the same way as in Bonvin et al. (2017) and Tewes et al. (2013a), who also give the procedure to derive the random and systematic errors from simulated light curves. As the results may depend on some of the key pa-

---

[1] PyCS can be obtained from `http://www.cosmograil.org`

rameters that characterize each method, we perform robustness checks identical to the ones in Bonvin et al. (2017); we use `PyCS` with a large range of method parameters, namely the number of knots of the spline technique and the covariance function of the *regression difference* technique (see Tewes et al. 2013a, for a full description of these parameters). We do not find significant differences in the mean time-delay values among the results obtained for the various tests, although the precision might vary.

The time-delay measurements are summarized in Fig. 4 for different data sets, i.e. with and without the Euler data. We note that the longer-baseline Euler data improve the time-delay estimates involving the D image when used in combination with the 2.2m WFI data. The latter have strong constraining power as high-frequency structures are captured in the curves, i.e. mostly the A and B components which display two strong features, including two inflection points in the case of B. For the much fainter D image, the situation is more complex as it displays only one shallow inflection point and no clear feature that can be matched to the other light curves.

3.2. Error estimates

In `PyCS`, the error estimates are done by running the curve-shifting techniques on mock light curves created from a generative model (see Tewes et al. 2013a). In these mocks, the intrinsic and extrinsic variations of the quasar are the same as the one infered from the real data as well as the temporal sampling and photometric errors. What changes from mock to mock are the correlated extrinsic variability (whose statistical properties mimic the observations), the photometric noise, the true time delays and the value of the simulated data points. The mock curves are drawn so that they have the same "time-delay constraining power" than the original data, i.e. the properties of the residuals after fitting the mock with a spline are statistically the same as on the real data.

We carry out simulations for a broad range of true input time delays around the measured value. This error analysis is summarized in Fig. 5 which, for each of the true time delay tested, provides the random and systematic errors. The final error for the delay is taken as the worst random error over all the bins, combined in quadrature with the worst systematic error. Obviously, the size of the bins, as well as the range of true time-delays explored when drawing the simulated light curves can have an impact on the final error. Part of the robustness checks we performed are intended to ensure that we do not overestimate or underestimate the errors by choosing unappropriate bin sizes and ranges in true time-delays. In the present case, our choice of possible true time-delays ranges up to ±10 days from our initial estimation obtained by running our point estimator on the original data. Such a wide range encompasses our uncertainty regarding the time delays of DES J0408−5354 that have never been measured before, yet is also small enough to make sure that simulated light curves (especially A and D) are sufficiently overlapping. We can note from the bottom panels of Fig. 5, that the large systematic errors of the spline optimizer (in blue and green) come mostly from the simulations with extremal values of true delays. However, the lack of clear and sharp variability structures in D, does not allow us to reduce this range of true time delays to be considered for the AD and BD delays.

We note that for the A and B light curves, which benefit from high-S/N data, the time delay is very accurate, whatever true delay is tested. This illustrates the importance of catching as many faint and short-duration structures in the light curves and the impact of high-cadence and high-S/N data. As a robustness test, we carry out the time-delay measurements without modeling explicitly extrinsic variations when using the spline technique. These results are indicated in black in Figs. 4 & 5 and show that the value of the time delays do not depend much on the extrinsic variations for the AB delay. Future observations of other objects will show if this is specific to DES J0408−5354 or a more general behaviour of the results with high-cadence and high-S/N light curves. Our final time delay value for AB is $\Delta t(AB) = -112.1 \pm 2.1$ days (1.8%), as obtained with the *free-knot spline* method using only the WFI data and extrinsic variations explicitly included. We make this choice because the time-delay measurement is precisely determined mostly thanks to the finely modeled peak in the WFI light curves (around mhjd = 57710 in A and 57820 in B). This peak is only crudely visible in the Euler data. Thus, adding the latter data set in the present case would only increase the overall noise. We also note that including extrinsic variations explicitly only slightly shifts the result while keeping the precision unchanged, as shown in Fig. 4. We chose nevertheless to explicitly include extrinsic variations since the residuals in the data speak in favor of it and since it is physically motivitated: microlensing and the subsequent extrinsic variations is present at some level in almost every lensed quasar know to date (Mosquera & Kochanek 2011).

In contrast to AB, the precision on the AD and BD delays depends on how extrinsic variations are modeled. Due to the lack of fine and sharp structures in the D light curve, and with only one shallow inflection point, the intrinsic and extrinsic variations are almost fully degenerate. We choose as our final results for these two delays the values obtained with the *free-knot spline* techniques for the combined Euler and WFI dataset. These times delay estimates are $\Delta t(AD) = -155.5 \pm 12.8$ days (8.2%) and $\Delta t(BD) = -42.4 \pm 17.6$ days (41%).

Finally, it is worth noting that the delays obtained with the *regression difference* technique applied only to WFI are consistent with the ones from the *free-knot spline* techniques, and the AD and BD delays are even more precisely measured. We however prefer to stick to the results of the *free-knot splines* since they are much more precise for AB, which is currently - and by far - the most constraining delay to be used in future modeling of this lensed system. We also explicitly avoid cherry-picking the best technique per delay, i.e. in the present case the *free-knot splines* for AB and the *regression difference* for AD and BD. Such an ad'hoc choice may introduce a bias difficult to quantify. Finally, the *free-knot splines* technique is the one giving the smallest systematics according to Fig. 5.

3.3. Comparison with simple models

Simple lens models are provided by Agnello et al. (2017) and are crucial to decide whether a new system will be useful for cosmology or not and to design the monitoring strategy. Thanks to these models, we knew that DES J0408−5354 would be challenging in terms of measuring time delays, as the longest one is half the visibility window of the object. Even under these difficult conditions, the MPIA 2.2m data allowed us to measure this long time delay to 1.8% accuracy and precision, by unveiling short and small photometric variations of the quasar images A and B. Our data also show that the extrinsic variations in the light curves do not affect the AB delay, as illustrated in the lower panels of Fig. 3.

The time delay we find for AB is in marginal agreement with the predictions of Agnello et al. (2017). However, at this stage of the modeling process, it would be hazardous to attempt any detailed comparison of our time delays with the model predictions,

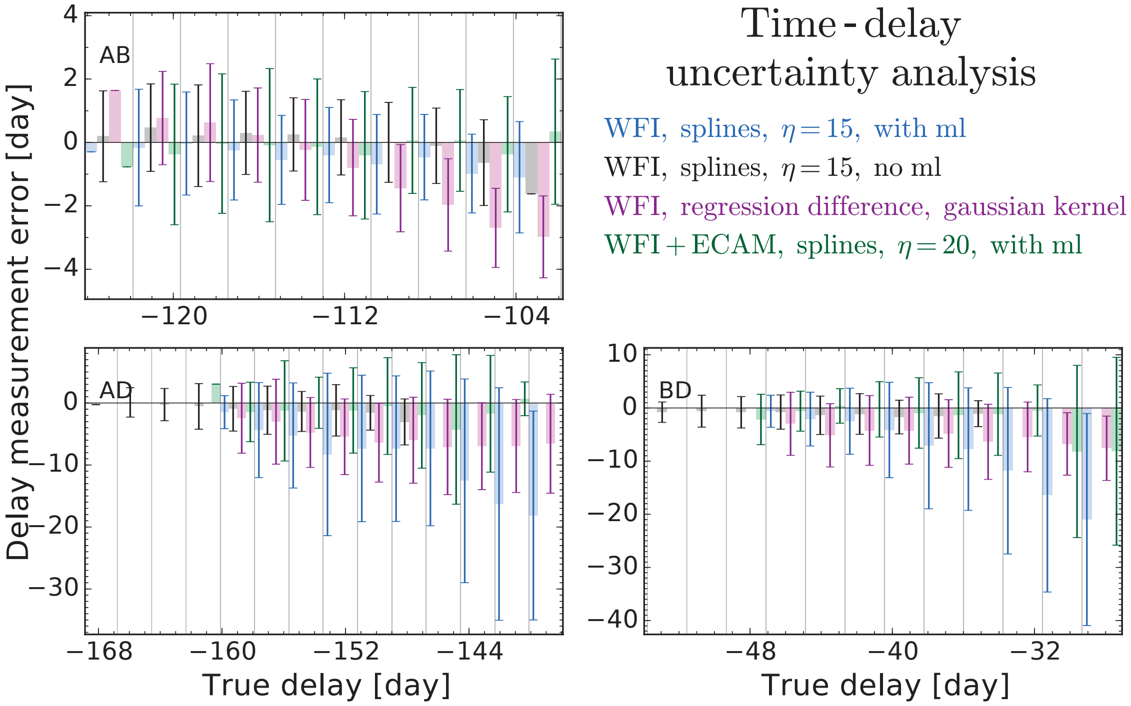


Fig. 5. Error estimates for the time-delay measurements performed on 1000 simulated light curves. The color code is the same as in Fig. 4. The x-axis of each panel shows the values for the true time-delay in the mock light curves. For each time-delay value are shown the random error bars as thin lines and the systematic errors as thick lines. The values for the time delays, as measured on the real data correspond to the center of each panel.

or to do cosmography with DES J0408−5354. This is not due to the quality of the modeling, but simply to the lack of deep and sharp imaging data and to the lack dynamical information on the lensing galaxy. Current data only provide the relative positions of the quasar images relative to the lensing galaxy and do not offer any information on the lensed host galaxy of the quasar. Deep HST imaging or ground-based AO is mandatory to do cosmography with any lens, as demonstrated by the H0LiCOW collaboration, and a velocity dispersion measurement for the lens is needed to lift some of the lensing degeneracies.

4. Conclusion

We demonstrate a new observational strategy for measuring time delays in lensed quasars, using high-cadence and high signal-to-noise monitoring photometry. The data, obtained almost daily over 7 months with the MPIA 2.2m telescope at ESO La Silla allow us to measure $\Delta t(AB) = -112.1 \pm 2.1$ days (1.8%), $\Delta t(AD) = -155.5 \pm 12.8$ days (8.2%) and $\Delta t(BD) = -42.4 \pm 17.6$ days (41%), where the error bars include systematics due to residual extrinsic variations. For the AB time delay, we note that the time-delay values depend little on the way extrinsic variations are modeled, hence indicating that the high-frequency signal in the light curves from the WFI instrument on the 2.2m telescope is dominated by the intrinsic variations of the quasar, as expected. For the D component, however, we only report a tentative delay due to the lack of fast variations seen in the light curve of this faint lensed image.

With the current imaging data for DES J0408−5354 it is too early to compare in detail the model predictions for the time delays (Agnello et al. 2017) and our measurements. Imaging with the HST or with ground-based adaptive optics (AO) is mandatory before drawing any conclusions and before doing any cosmological inference with DES J0408−5354. Fortunately, the object has several bright stars in its immediate vicinity, making it an excellent prey for VLT and AO, e.g. with the MUSE integral field spectrograph, providing the required dynamical information on the lensing galaxy(ies). In the near-IR observations with the VLT and the Hawk-I imager and the GRAAL AO system would both allow to measure the dynamics of the lens and probe the mass along the line of sight on a 7×7 arcmin field of view.

With the many wide-field surveys taking place at the moment (DES, KiDS, CFIS, DECals, HSC) and with LSST and Euclid coming in a fairly near future, the available number of lensed quasars will increase dramatically (e.g. Oguri & Marshall 2010). The present observations lend considerable hope to implement follow-up monitoring campaigns to measure reliable time delays quickly. We show how daily and very high signal-to-noise observations during 1 single season can match and potentially surpass long-term monitoring carried out at a lower rate (e.g. weekly) over many years. This should be accounted for when planning synoptic surveys such as the LSST. Although the latter will defenitely yield very high signal-to-noise images, the full benefit of monitoring data with a 8m telescope will be much enhanced in combination with a daily cadence or very close to a daily cadence.

We are currently monitoring 4 objects since October 2016 with the MPIA 2.2m. From a preliminary analysis of these 4 targets, we anticipate that reliable (i.e. to a few percents) time delays will be measured for 3 of them. We show here our results for DES J0408−5354, for which the observing season is finished. This demonstrate the effectiveness of high-S/N and high-cadence observations in measuring quickly and reliably time delays for many targets.

Acknowledgements. The authors would like to thank R. Gredel for his help in setting up the program at the MPIA 2.2m telescope. This work is supported by the Swiss National Science Foundation (SNSF). S.H. Suyu and D.C.Y. Chao thank the Max Planck Society for support through the Max Planck Research Group for SHS. T. Treu acknowledges support by the National Science Foundation through grant 1450141, by the Packard Foundation through a Packard Research Fellowship and by the UCLA Dean of Physical Sciences. K.Rojas is supported by Becas de Doctorado Nacional CONICYT 2017. T. Anguita and M. Chijani acknowledge support by proyecto FONDECYT 11130630 and by the Ministry for the Economy, Development, and Tourisms Programa Inicativa Científica Milenio through grant IC 12009, awarded to The Millennium Institute of Astrophysics (MAS). M. Tewes acknowledges support from the DFG grant Hi 1495/2-1. J. Garcia-Bellido is supported by the Research Project FPA2015-68048 [MINECO-FEDER], and the Centro de Excelencia Severo Ochoa Program SEV-2012-0249. C.D. Fassnacht acknowledges support from the National Science Foundation grant AST-1312329 and from the UC Davis Physics Department and Dean of Math and Physical Sciences.

Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey.

The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, the National Optical Astronomy Observatory, the University of Nottingham, The Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, Texas A&M University, and the OzDES Membership Consortium.

The DES data management system is supported by the National Science Foundation under Grant Number AST-1138766. The DES participants from Spanish institutions are partially supported by MINECO under grants AYA2015-71825, ESP2015-88861, FPA2015-68048, SEV-2012-0234, SEV-2012-0249, and MDM-2015-0509, some of which include ERDF funds from the European Union. IFAE is partially funded by the CERCA program of the Generalitat de Catalunya.

References

Agnello, A., Lin, H., Buckley-Geer, L., et al. 2017, ArXiv1702.00406
Agnello, A., Treu, T., Ostrovski, F., et al. 2015, MNRAS, 454, 1260
Bertin, E. & Arnouts, S. 1996, A&AS, 117, 393
Bonvin, V., Courbin, F., Suyu, S. H., et al. 2017, MNRAS, 465, 4914
Bonvin, V., Tewes, M., Courbin, F., et al. 2016, A&A, 585, A88
Cantale, N., Courbin, F., Tewes, M., Jablonka, P., & Meylan, G. 2016, A&A, 589, A81
Collett, T. E., Marshall, P. J., Auger, M. W., et al. 2013, MNRAS, 432, 679
Courbin, F., Chantry, V., Revaz, Y., et al. 2011, A&A, 536, A53
Courbin, F., Eigenbrod, A., Vuissoz, C., Meylan, G., & Magain, P. 2005, in IAU Symposium, Vol. 225, Gravitational Lensing Impact on Cosmology, ed. Y. Mellier & G. Meylan, 297–303
Eigenbrod, A., Courbin, F., Vuissoz, C., et al. 2005, A&A, 436, 25
Freedman, W. L. 2017, Nature Astronomy, 1, 0121
Freedman, W. L., Madore, B. F., Gibson, B. K., et al. 2001, ApJ, 553, 47
Hilbert, S., Hartlap, J., White, S. D. M., & Schneider, P. 2009, A&A, 499, 31
Jee, I., Komatsu, E., & Suyu, S. H. 2015, J. Cosmology Astropart. Phys., 11, 033
Liao, K., Treu, T., Marshall, P., et al. 2015, ApJ, 800, 11
Lin, H., Buckley-Geer, E., Agnello, A., et al. 2017, ApJ, 838, L15
Magain, P., Courbin, F., & Sohy, S. 1998, ApJ, 494, 472
McCully, C., Keeton, C. R., Wong, K. C., & Zabludoff, A. I. 2014, MNRAS, 443, 3631
McCully, C., Keeton, C. R., Wong, K. C., & Zabludoff, A. I. 2017, ApJ, 836, 141
Mosquera, A. M. & Kochanek, C. S. 2011, ApJ, 738, 96
Mushotzky, R. F., Edelson, R., Baumgartner, W., & Gandhi, P. 2011, ApJ, 743, L12
Oguri, M. & Marshall, P. J. 2010, MNRAS, 405, 2579
Ostrovski, F., McMahon, R. G., Connolly, A. J., et al. 2017, MNRAS, 465, 4325
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13
Rathna Kumar, S., Tewes, M., Stalin, C. S., et al. 2013, A&A, 557, A44
Refsdal, S. 1964, MNRAS, 128, 307
Riess, A. G., Macri, L. M., Hoffmann, S. L., et al. 2016, ApJ, 826, 56
Rusu, C. E., Fassnacht, C. D., Sluse, D., et al. 2017, MNRAS, 467, 4220
Sluse, D., Sonnenfeld, A., Rumbaugh, N., et al. 2016, ArXiv1607.00382
Suyu, S. H., Bonvin, V., Courbin, F., et al. 2017, MNRAS, 468, 2590
Suyu, S. H., Marshall, P. J., Auger, M. W., et al. 2010, ApJ, 711, 201
Suyu, S. H., Marshall, P. J., Blandford, R. D., et al. 2009, ApJ, 691, 277
Suyu, S. H., Marshall, P. J., Hobson, M. P., & Blandford, R. D. 2006, MNRAS, 371, 983
Tewes, M., Courbin, F., & Meylan, G. 2013a, A&A, 553, A120
Tewes, M., Courbin, F., Meylan, G., et al. 2013b, A&A, 556, A22
Treu, T. & Koopmans, L. V. E. 2002, MNRAS, 337, L6
Vuissoz, C., Courbin, F., Sluse, D., et al. 2008, A&A, 488, 481
Wong, K. C., Suyu, S. H., Auger, M. W., et al. 2017, MNRAS, 465, 4895

---

[1] Institute of Physics, Laboratory of Astrophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland

[2] Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL 60510, USA

[3] Department of Physics, University of California, Davis, CA 95616, USA

[4] Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA

[5] Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 452 Lomita Mall, Stanford, CA 94035, USA

[6] Max Planck Institute for Astrophysics, Karl-Schwarzschild-Strasse 1, D-85740 Garching, Germany

[7] Physik-Department, Technische Universität München, James-Franck-Straße 1, 85748 Garching, Germany

[8] Institute of Astronomy and Astrophysics, Academia Sinica, P.O. Box 23-141, Taipei 10617, Taiwan

[9] Department of Physics and Astronomy, University of California, Los Angeles, CA 90095, USA

[10] Departamento de Ciencias Físicas, Universidad Andres Bello Fernandez Concha 700, Las Condes, Santiago, Chile

[11] Millennium Institute of Astrophysics, Chile

[12] Instituto de Física y Astronomía, Universidad de Valparaíso, Avda. Gran Bretaña 1111, Playa Ancha, Valparaíso 2360102, Chile

[13] Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121, Bonn, Germany

[14] European Southern Observatory, Karl-Schwarzschild-Strasse 2, D-85748 Garching bei Munchen, Germany

[15] Centro de Astroingeniería, Facultad de Física, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, Macul 7820436, Santiago, Chile

[16] Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany

[17] Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, Casilla 603, La Serena, Chile

[18] Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

[19] Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

[20] LSST, 933 North Cherry Avenue, Tucson, AZ 85721, USA

[21] CNRS, UMR 7095, Institut d'Astrophysique de Paris, F-75014, Paris, France

[22] Department of Physics & Astronomy, University College London, Gower Street, London, WC1E 6BT, UK

[23] Sorbonne Universités, UPMC Univ Paris 06, UMR 7095, Institut d'Astrophysique de Paris, F-75014, Paris, France

[24] Kavli Institute for Particle Astrophysics & Cosmology, P. O. Box 2450, Stanford University, Stanford, CA 94305, USA

[25] SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA

[26] Laboratório Interinstitucional de e-Astronomia - LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil

[27] Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil

[28] Department of Astronomy, University of Illinois, 1002 W. Green Street, Urbana, IL 61801, USA

[29] National Center for Supercomputing Applications, 1205 West Clark St., Urbana, IL 61801, USA

[30] Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona) Spain

[31] Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA

[32] George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, and Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA

[33] Department of Physics, IIT Hyderabad, Kandi, Telangana 502285, India

[34] Institut de Ciències de l'Espai, IEEC-CSIC, Campus UAB, Carrer de Can Magrans, s/n, 08193 Bellaterra, Barcelona, Spain

[35] Instituto de Fisica Teorica UAM/CSIC, Universidad Autonoma de Madrid, 28049 Madrid, Spain

[36] Department of Astronomy, University of California, Berkeley, 501 Campbell Hall, Berkeley, CA 94720, USA

[37] Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

[38] Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA

[39] Department of Physics, The Ohio State University, Columbus, OH 43210, USA

[40] Astronomy Department, University of Washington, Box 351580, Seattle, WA 98195, USA

[41] Australian Astronomical Observatory, North Ryde, NSW 2113, Australia

[42] Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439, USA

[43] Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, CP 66318, São Paulo, SP, 05314-970, Brazil

[44] Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain

[45] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109, USA

[46] Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain

[47] Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA

[48] School of Physics and Astronomy, University of Southampton, Southampton, SO17 1BJ, UK

[49] Instituto de Física Gleb Wataghin, Universidade Estadual de Campinas, 13083-859, Campinas, SP, Brazil

[50] Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831

3.4.4 PG1115+080

PG1115+080 is a well known quasar whose time delays have been estimated six times already, yet based on only two different data sets. The first estimates were from a single monitoring season and from various telescopes, acquired in 1995/1996 and presented in Schechter et al. (1997) (hereafter Schechter data set). A reanalysis of these data followed shortly (Barkana, 1997; Pelt et al., 1998), confirming and refining the initial estimates. More than a decade later, Eulaers and Magain (2011) revisited the estimates with more recent curve-shifting techniques, and found the time-delay measurements being dependent on the technique employed. In parallel to this, three seasons of observations with the Maidanak telescope between 2004 and 2006 (hereafter Maidanak data set) were presented in Vakulik et al. (2009), along with new estimates of the time delays significantly different from the previous publications. These new estimates were then revised by Tsvetkova et al. (2016), but also processed independently by Shimanovskaya et al. (2015) who found results in better agreement with the one published from the Schechter data set. Figure 3.14 presents these six sets of published time-delay estimates.

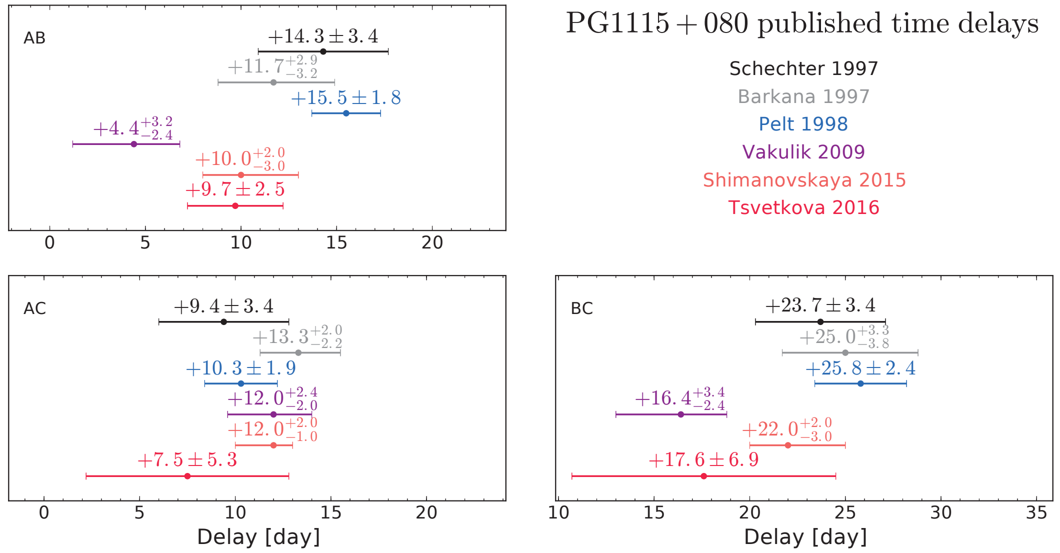


Figure 3.14: Published time-delay estimates of PG1115+080. The Schechter et al. (1997); Barkana (1997); Pelt et al. (1998) estimates make use of the Schechter data set, whereas the Vakulik et al. (2009); Shimanovskaya et al. (2015); Tsvetkova et al. (2016) estimates make use of the Maidanak data set. Error bars represent the 1σ confidence interval.

The difference between the estimates on the Schechter and Maidanak data set can be explained by the fact that the two data sets are short and do not cover the same period of time, but also by the different curve-shifting techniques used. In order to lift the uncertainty on these discrepancies, we decided to apply the same PyCS formalism on the three data sets available. The Schechter data have been graciously handed over to us by P. Schechter, the Maidanak R-band data are publicly available from Tsvetkova et al. (2010) and we have our own WFI data presented in Sec. 2.5.5. Figure 3.15 presents these two data sets, along with the WFI data already presented in Sec. 2.5.5 and reproduced here for convenience.

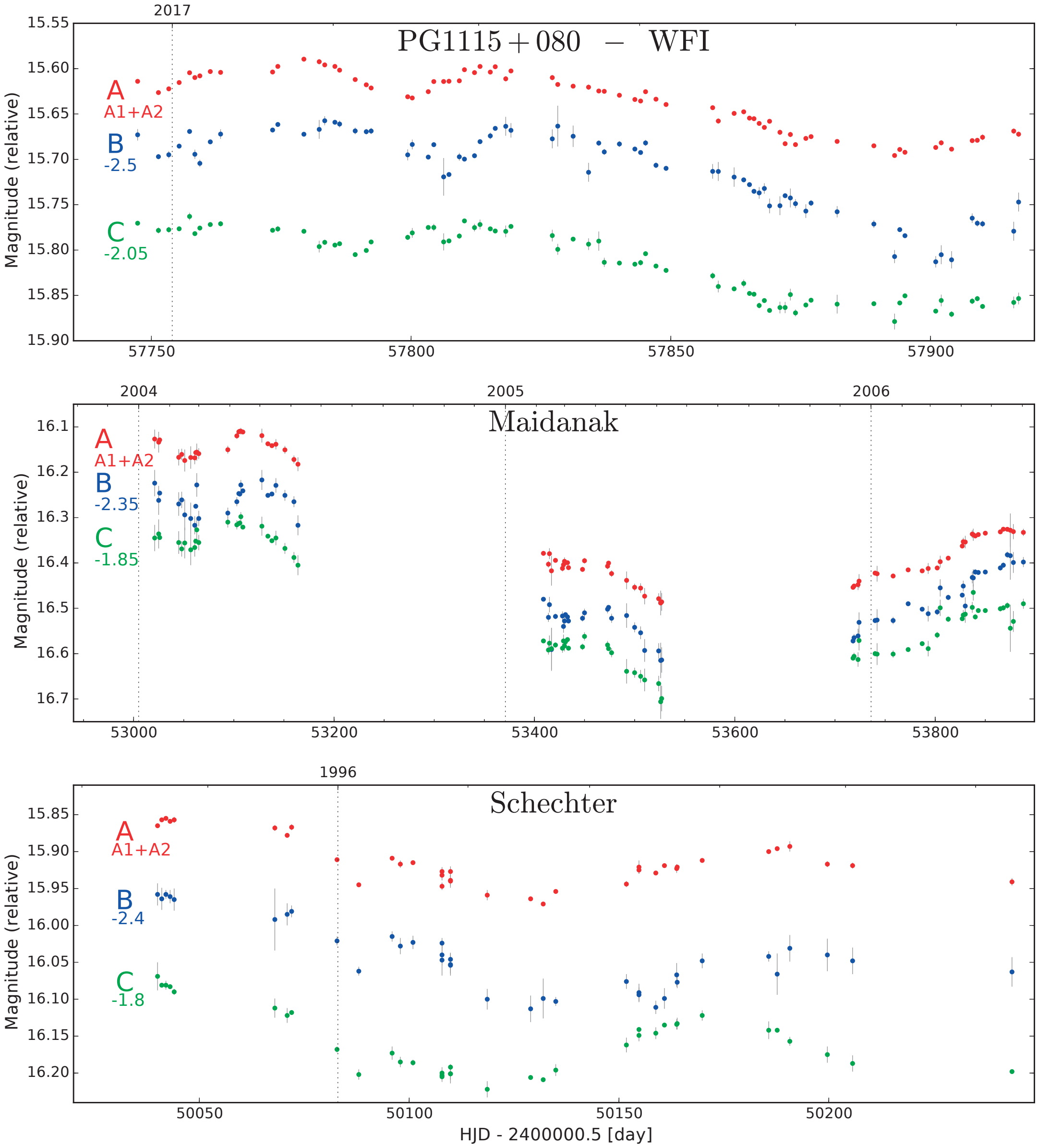


Figure 3.15: Light curves of the PG1115+080 data sets; one season from WFI in 2017 (top), three seasons from Maidanak in 2004-2006 (Tsvetkova et al., 2010, middle) and one season from Schechter in 1995/1996 (Schechter et al., 1997, bottom). The light curves of each data sets have been shifted in magnitude for visual purposes. The shifts are indicated in the corresponding panels. The error bars on the WFI and Maidanak data sets correspond to the photometric error bars, whereas on the Schechter data set they correspond to the standard deviation between multiple measurements taken during the same night of observation.

We apply the free-knot splines technique formalism (see Sec. 3.2.1 and 3.2.3) independently on these three sets of data, exploring in each case the number of knots η of the spline modeling the intrinsic variations. The extrinsic variability is also explicitly modeled using splines; when there is only one monitoring season (Schechter and WFI data), we use only three knots (two at the boundaries and one in the middle of the extrinsic variability curves) so that the number of spline pieces is $n_{ml} = 2$, similarly to the analysis of DESJ0408-5354. For the three seasons of the Maidanak data set we use an initial knot step of $\eta_{ml} = 200$ days, and the knots are free to move in a window around their initial position similarly to the analysis of HE0435-1223. Figure 3.16 presents a blind comparison of our results on the three data sets.

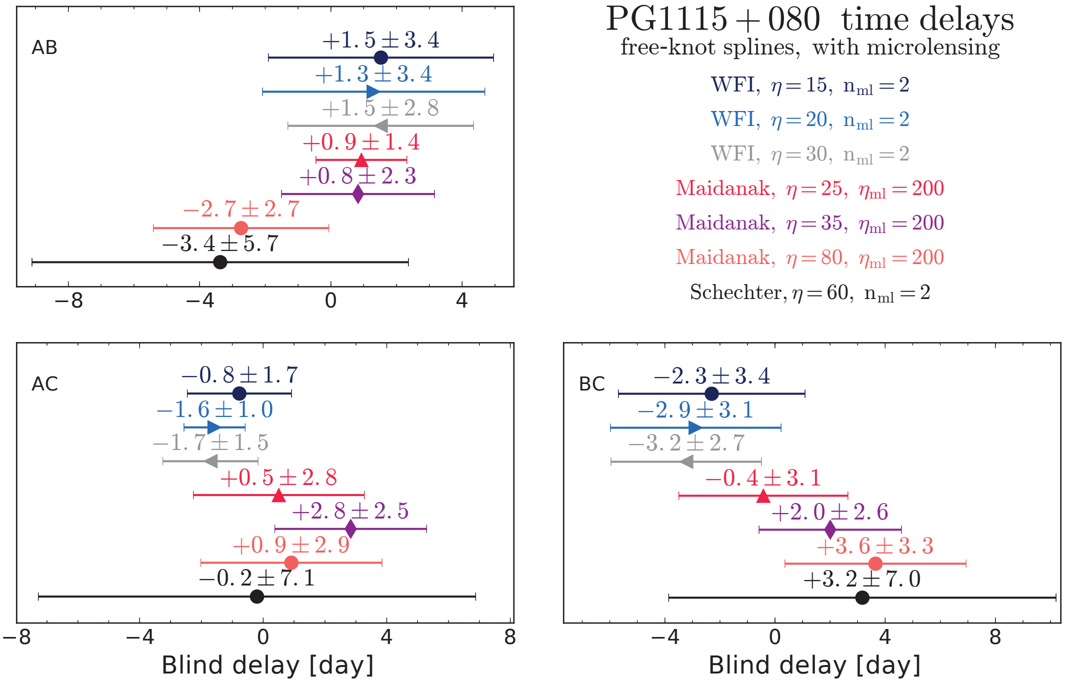


Figure 3.16: Blind time-delay estimates of the lensed quasar PG1115+080 from the Schechter, Maidanak and WFI data sets. Each set has been individually analyzed using the free-knot splines technique. η refers to the initial knot step of the intrinsic variability curve, η_{ml} to the initial knot step of the extrinsic variability curves (Maidanak only) and $n_{ml} = 2$ to the number of spline pieces forming the extrinsic variability curve (WFI and Schechter only).

Here, blindness is enforced to avoid unconsciously fine-tuning the method parameters of the free-knot spline technique in order to reproduce (or not) some of the results previously published on the Schechter and Maidanak data sets. Indeed, we note that both the time delays and associated uncertainties of the Maidanak data are sensitive to the initial knot step η. Although not significantly in tension with each other, a change of e.g. 4 days in the BC delay is large enough to significantly impact the inferred time-delay distance and subsequent H_0 measurement. Whether these variations can be corrected for, partly disregarded or have to be considered seriously is still an open question. On the other hand, we note that the time delays measured on the WFI data are more robust against a change in the method parameters, although the uncertainties are not. Results from the Schechter data are much less precise than their Maidanak and WFI counterpart, as expected from only one season with a weekly monitoring cadence.

The PG1115+080 time delays are going to be used in combination with ground-based Adaptive Optics imaging of the lens galaxy to infer its time-delay distance (C. Fassnacht, private comm.). It is thus of a crucial importance to determine a precise and accurate value of the time delays. If the results of the free-knot splines technique on the three data sets are still in agreement, the question that arises is how other curve-shifting techniques perform on these data sets, and then how to combine the various estimates and data sets in order to have only one time-delay estimate per pair of quasar images. It would be definitely dishonest to disregard one of the data sets unless we can demonstrate that a change in the method parameters - or a change of curve-shifting technique - yield inconsistent results. However, a data set yielding consistent time delays (like WFI) should have more weight on the final combination than one that does not (like Maidanak). All of these being currently work in progress, this question will hopefully be settled in a very near future.

3.4.5 RXJ1131-1231, WFI2033-4723 & HE1104-1805

These three lenses are regrouped in a single subsection for a reason; they are the currently "hot" targets on which the H0LiCOW collaboration is currently working on. For that reason, there is not much I can say about them ahead of the official publications. At the time this document is written, there is still much to be done but the work is off to a good start. In any case, no results are presented in this section; I instead briefly recall the existing work and give near-future forecasts regarding the time-delay measurements of these three objects.

The most recent COSMOGRAIL time-delay measurements for RXJ1131-1231 are presented in Tewes et al. (2013b). Interestingly enough, Morgan et al. (2006) present their own time-delay measurements of RXJ1131-1231 on a much shorter monitoring campaign of two years only, using similar facilities and curve-shifting techniques than in the analysis of HE0435-1223 performed by Kochanek et al. (2006). This time, the COSMOGRAIL estimates are more precise but also sensibly different from the Morgan et al. (2006) ones that seemed to suffer from unacknowledged presence of microlensing. The values of Tewes et al. (2013b) have already been used in combination with high-quality follow-up data to infer H_0 (Suyu et al., 2014; Birrer et al., 2016; Bonvin et al., 2017), but recent improvements on the lens modeling technique as well as a few extra monitoring seasons will together increase the precision of the time-delay distance measured from this system.

WFI2033-4723 and HE1104-1805 are the next two targets that are going to be fully analyzed by the H0LiCOW collaboration. The most recent COSMOGRAIL estimate of the time delays of WFI2033-4723 are presented in Vuissoz et al. (2008) on only three years of data, along with an estimate of H_0 using simple lens modeling. Regarding HE1104-1805, there are no public COSMOGRAIL results so far. The most recent estimates were proposed by Poindexter et al. (2007) that proposed a delay of ~ 152 days, confirming a previous estimate from Ofek and

Maoz (2003). However, earlier works by Wisotzki et al. (1998); Gil-Merino et al. (2002) found a time delay of ~ 300 days, although using much poorly sampled data. For both WFI2033-4723 and HE1104-1805, a wealth of new data has been acquired since their respective most recent publications. The forecast in terms of time-delay measurements and, consequently, H_0 estimate are optimistic. With new data from the Euler and 2m2 telescopes as well as from the SMARTS consortium, we expect a precision on the time-delay measurements of 4%, respectively 2% on WFI2033-4723 and HE1104-1805 (Suyu et al., 2016).

4 Measuring the Hubble constant at 3.8% precision

> Nothing compares to the measurement of the Hubble constant
> in bringing out the worst in astronomers.
>
> *- Christopher S. Kochanek -*

In this last chapter, I present the culmination of everything developed earlier. Designing monitoring campaigns of lensed quasars and measuring the resulting time delays with great precision and accuracy served one main purpose, which is measuring the Hubble constant H_0. Well sampled light curves of lensed quasars can of course be used in other contexts as well, such as the study of the lensed galaxy properties through microlensing or the study of quasar variability (see Sec. 1.2.4). Yet, measuring H_0 has always been the primary focus of COSMOGRAIL.

In the following, I start by recalling the current context in which the most recent H0LiCOW measurement folds in. Doing so, I briefly review the most recent values of H_0 published since the turn of the century. Then, I present the results of the H0LiCOW collaboration, notably detailing how the lens modeling and line-of-sight contribution are performed. I finally reflect on the implication of our results when compared to the most up-to-date values of H_0 measured by other direct and indirect techniques. I conclude by forecasting the future of the H0LiCOW collaboration and the time-delay cosmography in general, in regard of what has been achieved so far.

4.1 Measuring the Hubble constant

Before describing the current cosmological context in which my work is encompassed, let us have a brief recap of the history of H_0 measurements. A very peculiar property of the Hubble constant is that astronomers working on its measurement seem to have a hard time agreeing on what its value should be. The initial measurement by Lemaître was already quite controversial, with a value of $H_0 \simeq 625\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Lemaître, 1927): such an expansion rate, once inverted, led to an age of the Universe of approximately two billion years. This was in direct contradiction with radioactive dating of the Earth's components, indicating already in the 1930's an age of at least three billion years for our planet. That controversy was solved in the late 1950's where combined progresses in the understanding of physics of variable stars allowed to significantly improve their period-luminosity calibration, thus leading towards a value of $H_0 \simeq 75\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Sandage, 1958). But then, two different solutions seemed to emerge. By the 1970's, Sandage and Tammann (1975) advocated for $H_0 \simeq 55\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ were de Vaucouleurs and Bollinger (1979), among others, claimed towards a higher value of $H_0 \simeq 100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. That controversy, reportedly bitter, was exacerbated by the fact that the two opposing groups were using the same technique based on standard candles. Other techniques to measure H_0 emerged during this period, notably time-delay cosmography. Yet, they were neither able to provide a consensus on their own (see e.g. Sec. 1.2.3). It is finally decades later, at the dawn of the 21st century that this long-lasting controversy was solved with the construction and operation of the Hubble Space Telescope. One of its Key Projects consisted on measuring extra-galactic distances to an unprecedented precision, hence determining the expansion rate of the Universe at 10% precision. At the turn of the millennium, after a decade of observations, Freedman et al. (2001) presented a value of $H_0 = 72{\pm}8\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. And finally, in the decade that followed, no more controversies showed up. Wendy Freedman, one of the scientists at the core of the Hubble Space Telescope measurement declared in a recent report on the issue *"I never thought I'd work on the Hubble constant again"* (Sokol, 2017). The issue seemed settled. But what happened in the past was to happen again, only this time on a different scale, and possibly with a different conclusion.

4.1.1 H_0 in the Standard Cosmological Model

1.3

As stated in the introduction of this manuscript, the idea of trying to understand and synthesize the Universe we are living in is certainly as old as the idea of the Universe itself, and all the scientific models that emerged over the ages had to adapt to the observations. The Big Bang cosmological model, towards which Lemaître cleverly hinted in 1927 and that is firmly established since the discovery of the CMB by Penzias & Wilson in 1964 (see Sec. 1.1.7) had to be adapted following the discovery of the accelerated expansion of the Universe by Riess et al. (1998); Perlmutter et al. (1999). Thus, the early 2000's saw the emergence of a golden standard in cosmology, an analogous to the standard model in particle physics but for the Universe: the *Standard Cosmological Model*, or flat-ΛCDM. It consists of an Universe supposedly without spatial curvature, containing dark energy acting as a cosmological constant denoted Λ and slowly-moving and weakly/non-interacting exotic matter referred to as cold dark matter, or CDM. The success of the ΛCDM model comes from its relative simplicity - only six physical parameters are needed to fully describe the Universe's content and current observations. Remarkably, the ΛCDM model accorded very well with the exponentially growing number of cosmological and astrophysical experiments and observations, or at least with the vast majority of them.

The Hubble constant is not one of the six base parameters of the ΛCDM model, but can be *predicted* once the base parameters have been determined. Following the introduction of the ΛCDM model, Cosmic Microwave Background experiments like WMAP independently refined their prediction of the Hubble constant value, incrementally increasing their overall precision after each year of operation. With nine years of data reduced into a single map, (Bennett et al., 2013b) presented a value of $H_0 = 70.0 \pm 2.2\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, still in agreement with Freedman's prediction established a decade earlier. But at the same time, a discordant voice started to be heard, and not the least. Adam Riess, a few months before being awarded his Nobel prize revised the distance ladder estimate, using again the Hubble Space Telescope to $H_0 = 73.8 \pm 2.4\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Riess et al., 2011). That value was in itself not significantly different from the WMAP prediction, but when the WMAP data were combined with finer-scale CMB measurement from ground-based telescopes (the Atacama Cosmology Telescope and South Pole Telescope), Baryon Acoustic Oscillations and the direct H_0 measurement from (Riess et al., 2011) itself, it yielded a value of $H_0 = 68.3 \pm 0.8\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, in tension with the distance-ladder estimate alone. Furthermore, in the same year that saw the publication of the final WMAP results, the first results from the Planck satellite were also published. Planck Collaboration et al. (2014) proposed a value of $H_0 = 67.3 \pm 1.2\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ using CMB temperature observations from Planck alone. The distance ladder technique using the Tully-Fisher relation of SNIa host galaxies as intermediate calibrator yielded $H_0 = 75.2 \pm 3.0\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Sorce et al., 2012), and a re-calibration of the Cepheids rung yielded $H_0 = 74.3 \pm 2.1\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Freedman et al., 2012). If the tension between the various distance ladder and WMAP results could have been at least partially explained through the combination of data sets of the former, that could not be the case anymore regarding the tension with Planck. But the distance ladder measurement were, however, nor bias-free neither completely in control of the sources of systematic errors (see e.g. Efstathiou, 2014; Rigault et al., 2015; Zhang et al., 2017). Planck results being in an extremely good agreement with the ΛCDM model, skepticism over the distance-ladder results prevailed and the general view leaned towards a concordance value for H_0 that, sooner or later, should emerge (Bennett et al., 2014).

In 2015, the Planck team published a new iteration of their results, using the full-mission temperature and light polarization map acquired from 2009 to 2013. Similarly to the 2013 intermediate data release, the full-mission data release found no deviation from the ΛCDM model. Although potential inconsistencies were pointed out (e.g. Addison et al., 2016), the overall precision of the measurements improved as well, with a predicted value of the Hubble constant of $H_0 = 67.8 \pm 0.9\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Planck Collaboration et al., 2016c). Combined with the most up-to-date Baryon Acoustic Oscillation survey, Alam et al. (2016) found $H_0 = 67.6 \pm 0.5\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Similarly, Doux et al. (2017) found $H_0 = 68.7 \pm 0.9\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ by combining Planck and large-scale structures. The constraints from Planck alone were tightened even further in a more recent analysis that corrects from systematic effects in the polarization maps, yielding $H_0 = 66.93 \pm 0.62\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Planck Collaboration et al., 2016d). And then the distance ladder technique stroke back. With new data acquired since 2011 and an improved analysis taking into account some of the criticism previously addressed, a revised value of H_0 $=73.24\pm1.74\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ was published (Riess et al., 2016), at a 3.3σ tension with Planck's results. In summary, both group revised their work with more data and a better control over the possible sources of error, only to confirm and strengthen their disagreement.

4.1.2 Adding time-delay cosmography

Despite a strengthening tension between the distance ladder measurement and CMB prediction of H_0, the possibility that new physics beyond the standard cosmological model could arise is only warily mentioned, and this certainly rightly. The trending expression is more to speak about *unknown unknowns*, i.e. sources of bias of unknown nature affecting the measurements and that are currently unaccounted for, since invisible in the scope of each individual experiment. Either way, clarifying the situation desperately requires a referee, that takes the form of independent, complementary cosmological probes. X-rays observations of clusters of galaxies coupled to measurement of the Sunyaev–Zel'dovich effect as proposed by Bonamente et al. (2006) yielded $H_0 = 76.9^{+13.9}_{-11.9}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Using water masers emission, Gao et al. (2016) found $H_0 = 66.0\pm6.0\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Observing extra-galactic regions of ionized hydrogen, Chávez et al. (2012) found $H_0 = 74.3\pm6.0\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. All these techniques were promising but not precise enough (yet?) to bring something new in the discussion around the CMB versus distance ladder tension.

Time-delay cosmography, after having initially wandered between unsubstantial values (see Sec. 1.2.3) had a second breath at the turn of the millennium when precise and high-cadence monitoring campaigns of lensed quasars such as COSMOGRAIL started to produce light curves and time delays of much better quality than before. Yet, new controversies inside the field emerged. The paucity of information regarding the lens systems and their surroundings forced the investigators to make crude assumptions during the lens modeling. As a result, the published values were still exhibiting inconsistencies depending on what these assumptions were. A stringent illustration of this issue is for example the work from York et al. (2005) that proposed two inconsistent values for the Hubble constant depending on whether the spiral arms of the lens galaxy were accounted for in the lens modeling ($H_0 = 70.0\pm5.0\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ at 2σ) or not ($H_0 = 61.0\pm7.0\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ at 2σ). Similarly, Vuissoz et al. (2007) noted a very strong dependence of the value of H_0 according to the modeling assumption, for example $H_0 = 80.8^{+7.0}_{-3.0}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ if the lens galaxy follows a De Vaucouleurs mass profile versus H_0 $= 51.7^{+4.0}_{-3.0}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ for a singular isothermal sphere model. Using a singular isothermal ellipsoid model on a single lens system, Paraficz et al. (2009) found $H_0 = 73.0^{+3.0}_{-4.0}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ whereas combining five systems together yielded $H_0 = 61.5^{+8.0}_{-4.0}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. However, joint analysis of larger samples of lenses considered altogether were in a relatively better agreement, providing results of similar accuracy. Using 16 systems, Oguri (2007) found $H_0 =$ $68.0\pm14.0\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ where Coles (2008) found $H_0 = 71.0^{+6.0}_{-8.0}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ with 11 systems, both measurements being in good agreement with other methods.

The breakthrough came with the identification and acknowledgment of the main issues that plagued the previous individual measurements. Not only was a precise lens modeling crucially important, but so were the environmental effects due to all the galaxies along the line-of-sight. With in hand data of unprecedented quality taken by the Hubble Space Telescope, Suyu et al. (2010) found $H_0 = 70.6 \pm 3.1\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (although fixing the density ratio of matter in the present universe at Ω_m=0.3) from the lens system B1608+656. Similarly, on the lens system RXJ1131-1231, Suyu et al. (2014) found $H_0 = 80.0^{+4.5}_{-4.7}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ under similar conditions. Although the values of H_0 from these two systems are still in tension. It can be explained by partially known effects unaccounted for (see e.g. Sec. 4.2.2), unknown unknowns, a statistical fluke or even the effect of new physics. In any case, the confidence in these two recent measurements has been drastically improved to the point is that any of the option mentioned above can be fairly considered.

Built on the individual measurements from B1608+656 and RXJ1131-1231, the H0LiCOW collaboration published its analysis of a third lensed system, HE0435-1223 and regrouped the three measurements into a single one that can be independently expressed in any underlying cosmological model. Hence, Bonvin et al. (2017) found a value of the Hubble constant of $H_0 = 71.9^{+2.4}_{-3.0}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ in a ΛCDM model. This value is in moderate tension with the most recent measurement from Planck, yet in a very good agreement with the distance ladder. Note that by revising the prior assumptions, H0LiCOW found a slightly higher value (unpublished) of $H_0 = 72.5^{+2.4}_{-2.4}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

Figure 4.1 presents a comparison between the different values of H_0 found by the most recent measurements of WMAP, Planck, the distance ladder and time-delay cosmography. Since the latter two estimates are established in a completely independent manner, they can be combined in a single *local* probe, in opposition to the *distant* probe from the CMB, and thus increasing further the tension between the local and distant probes. Let us note that a fair local versus global comparison would imply to combine the results from the Sunyaev-Zel'dovich clusters from Bonamente et al. (2006), the water masers from Gao et al. (2016) and the HII galaxies from Chávez et al. (2012) into the local probe, as well as rescaling the distance ladder results of Riess et al. (2016) according to the other possible calibrators as e.g. Sorce et al. (2012); Freedman et al. (2012), which has not been done here - the goal being mostly to illustrate the stringency of the tension than to precisely quantify it.

So much for the current state of play. Obviously, the addition of the H_0 measurement from strong lensing is not enough to discriminate between any possible scenario. However, it renders more plausible the option that new physics not considered so far is behind the discrepancy between the high- and low-redshift observations. As noted in recent reports (e.g. Sokol, 2017; Freedman, 2017), cosmology is now at a crossroad, yet a very foggy one. Currently, the best way to clear up the path is to acquire more data. Planck will publish in a near future a final analysis of its maps, the distance ladder technique is coming up with new calibrators (Beaton et al., 2016; Hatt et al., 2017) and a bunch of new strong gravitational lenses are being discovered and thoroughly analyzed (e.g. Suyu et al., 2016; Agnello et al., 2017; Courbin et al.,

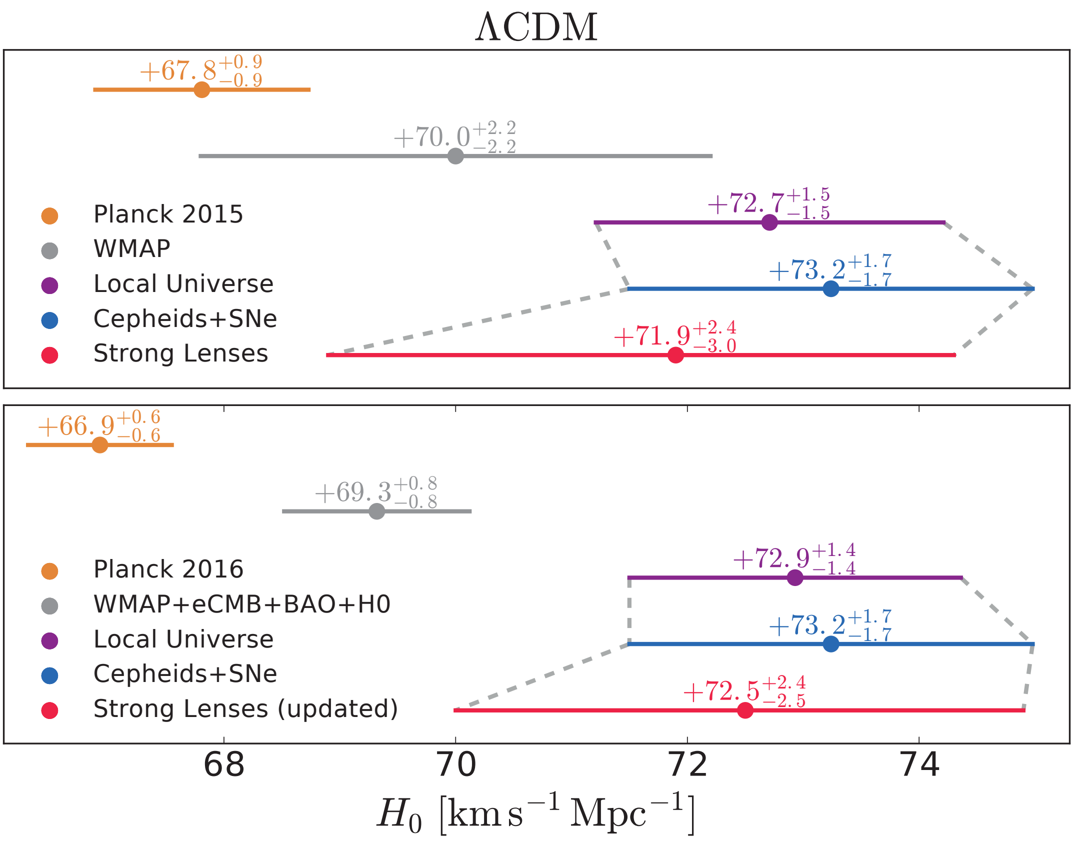


Figure 4.1: Comparison of the most recent H_0 values. The CMB predictions from Planck (Planck Collaboration et al., 2016c) (top panel) and its updated value (Planck Collaboration et al., 2016d) (bottom panel) are in brown. The predictions from WMAP alone (top panel) or combined with finer scales measurements from SPT and ACT, BAO and using a prior from (Riess et al., 2011) are in gray (Bennett et al., 2013b). The local Universe combined measurement are in purple. They result of the combination of the distance ladder technique using Cepheids and Type-Ia Supernovae in blue (Riess et al., 2016) with the time-delay cosmography (Bonvin et al., 2017) (top panel) and its updated value (bottom panel) in red.

2017). Whether this is going to be decisive or not is still to be discovered.

4.2 The H0LiCOW collaboration

H0LiCOW is an international collaboration that regroups ~30 scientists in various institutions around the world. It stands for H_0 *Lenses in COSMOGRAIL's Wellspring*, an acronym built around another acronym. Written this way, H0LiCOW acknowledges the central importance of the time-delay measurement in the determination of the Hubble constant, yet highlights the fact that there are other significant steps that need to be carefully executed to yield a proper H_0 measurement. Plus, it sounds extremely catchy, which matters probably more than it should.

The H0LiCOW collaboration has published seven papers to date, an eighth one being currently in preparation. The first paper (Suyu et al., 2016) presents the project and forecasts for the near future of time-delay cosmography. The second to fifth papers (Sluse et al., 2016; Rusu et al., 2017; Wong et al., 2017; Bonvin et al., 2017) focus on the analysis of HE0435-1223 and its combination with previous results on B1608+656 and RXJ1131-1231, in order to determine a joint value of H_0. The sixth and seventh papers (Ding et al., 2017a,b) make use of the previous results to constrain the correlation between black hole mass and host galaxy luminosity. In this section, I present in details how the H0LiCOW most recent measurement of H_0 is computed. Since I already dedicated the previous chapters of this thesis to time-delay measurements, it is only fair that I now focus more on the other aspects of time-delay cosmography, namely the mass modeling and line-of-sight mass contribution. I also detail how the measurements from the three strongly lensed quasars have been combined and why that combination makes sense, and present the kind of results we get when combining time-delay cosmography with other cosmological probes. The H0LiCOW results were announced in a press release that aired in January 2017 and made the headlines of both mainstream newspapers and scientific-oriented publications and institutions. It is described in further details in Appendix B.

4.2.1 On the importance of a blind analysis

Each time it was necessary, our intermediate measurements and subsequent analysis where carried out blindly. This has been done in order to avoid as much as possible to be affected by a *confirmation bias*. In practice, knowing e.g. the time delays or the lens galaxy velocity dispersion without any prior knowledge on the modeling intermediate results cannot really bias the investigators, as the link between these measurements and the cosmology is complex and strongly degenerated with other parameters of the whole analysis. On the other hand, all the modeling efforts performed in H0LiCOW are carried out without any knowledge of what the cosmological outputs are. For example, the time-delay distance determination presented in Fig. 4.3 a little bit further has been shifted by a constant blind offset, since it directly relates to cosmology through Eq.2.11. It is only a few weeks before the publication of the results, once all the necessary robustness tests at each step of the analysis have been carried out thoroughly that the truth is unveiled and the cosmological parameters computed. The measurement of H_0 on HE0435-1223 enforced strictly that rule, and no modifications were allowed after the unblinding. Such a rule was also applied for the first cosmological inference computed on RXJ1131-1231 by Suyu et al. (2013), although the re-analysis using new mass profiles in Suyu et al. (2014) that was then used in Bonvin et al. (2017) was consequently not blind anymore.

By enforcing blindness in the H0LiCOW collaboration, we also hope to establish a framework whose integrity cannot be doubted. The idea is in place since many years already in other fields of Physics (e.g. in Particle Physics), and the benefits of implementing it in Cosmology have been already recognized (e.g. Croft and Dailey, 2011; Maccoun and Perlmutter, 2015). Recently, Zhang et al. (2017) developed such a framework for the distance ladder measurement technique using Cepheids and Supernovae, and DES Collaboration et al. (2017) did the same with 2-point correlation function analysis of galaxies in the DES survey. We are, however, well conscious that achieving total blindness can be extremely challenging. In the upcoming years, when the full initial H0LiCOW sample of five lenses will be jointly analyzed, results of previously analyzed lenses will be updated with new data and improved techniques. This will require an extra dose of cautiousness; for example, new time-delay estimates yielding shorter time delays would favor a larger value of H_0 than what has been previously measured. It will thus be very important at that time to have a proper implementation of the blindness in `PyCS`.

4.2.2 Modeling the lens

The lens modeling refers to the act of constructing and constraining a numerical model of the lens galaxy that is able to reproduce all the observables (time delays, lensed images positions and flux ratios, lens galaxy light profile, extended arc-like structures, etc...) while respecting the time-delay distance equation (Eq. 2.11), reproduced here for convenience.

$$\Delta t_{ij} = \frac{D_{\Delta t}}{c}\left[\frac{(\boldsymbol{\theta}_i - \boldsymbol{\beta})^2}{2} - \psi(\boldsymbol{\theta}_i) - \frac{(\boldsymbol{\theta}_j - \boldsymbol{\beta})^2}{2} + \psi(\boldsymbol{\theta}_j)\right]. \tag{4.1}$$

The Δt_{ij} term is the time delay between two lensed images i and j, $D_{\Delta t}$ is the so-called time-delay distance, a quantity that depends on the redshift of the lens and source as well as on H_0, $\boldsymbol{\theta}_i$ is the angular position of the lensed image i on the plane of the sky, $\boldsymbol{\beta}$ is the inferred angular position of the source, and ψ is the gravitational potential of the lens.

By reproducing as many observational constraints as possible, the lens modeling can reconstruct the source position $\boldsymbol{\beta}$, and thus constrain the time-delay distance $D_{\Delta t}$ through Eq. 4.1. That would work perfectly well in the case of an isolated lens galaxy of known mass distribution, all alone in the Universe if not for the observer and the source. However, reality is more complex: the deep sky is a crowded place and other galaxies in the line-of-sight, close to the lens galaxy in angular projection but not necessarily in absolute distance can significantly affect the lensing observables, and must thus be properly folded into the lens modeling. We distinguish two regimes; when another perturbing galaxy is close enough so that it must be explicitly included in the modeling as a secondary lens, or when it can be considered only as an external perturbation averaged over all the other galaxies along the line-of-sight. The former is discussed further below in this section, whereas the latter is discussed in Sec. 4.2.3. Figure 4.2 illustrates this with in the case of HE0435-1223, presenting a close-up view on the environment around the lens.

What makes the lens modeling really tricky is that there is a degeneracy between the choice of the lensing potential of the main lens and the product of the time delay and the Hubble constant $\Delta_t H_0$, that leaves all the other lensing observables unmodified. It means that without observational constraints on the lens galaxy other than the time delays, lensed images position

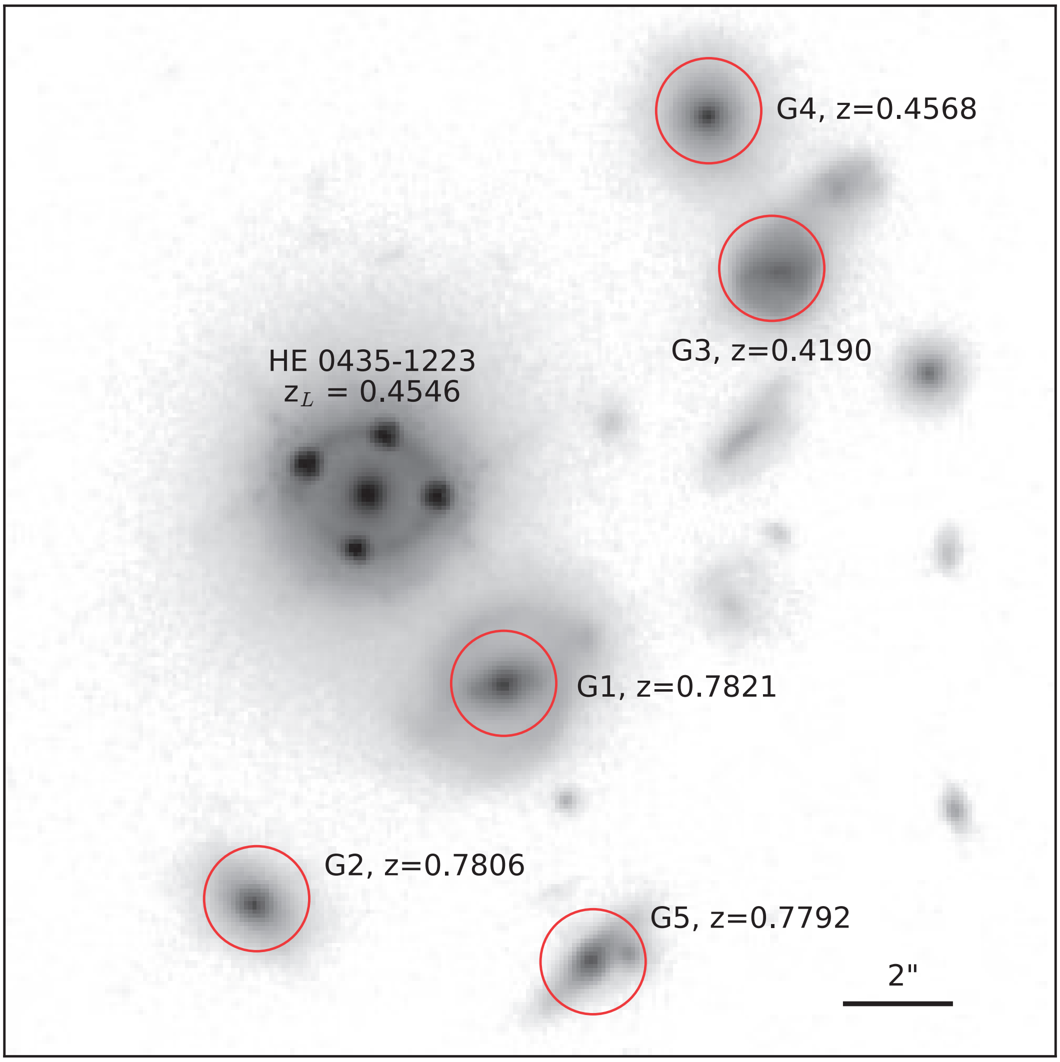


Figure 4.2: HST image in the F160W-band of a 20x20 arcseconds field of view of the lensed quasar HE0435-1223. The five most significant perturbing galaxies are labeled G1 to G5 and circled in red. The annotated redshifts z were measured using various ground-based spectrographs - see Sec. 3.4 of Wong et al. (2017) for more details. Adapted from Wong et al. (2017).

and flux ratio, various lensing potential could reproduce the observations, yielding in each case a different H_0! That degeneracy known today as the *mass-sheet transformation* was first introduced by Falco et al. (1985) and further developed by many authors (see e.g. Liesenborgs and De Rijcke, 2012; Schneider and Sluse, 2013, and references therein). Xu et al. (2016); Tagore et al. (2017) explored from numerical simulations how such an effect, if unaccounted for would shift the time-delay distance measurement and the Hubble constant up to several dozens of percents. Recently, Schneider and Sluse (2014); Unruh et al. (2017) introduced a generalization of the mass-sheet transformation called the *source-position transformation* that extends the range of possible degeneracies; whereas the mass-sheet transformation is a mathematically exact degeneracy, the source-position transformation is not necessarily exact, but the current modeling techniques are not precise enough to properly discriminate between two solutions. According to these authors, it is plausible that the majority of the lens-to-lens discrepancy in the inferred H_0 comes from a source-position transformation not correctly accounted for.

In such a framework, the observational constraints need not only to be as precise as possible in order to constrain at best Eq. 4.1, but also to break the degeneracies mentioned above. In terms of modeling, the time delays are one of the necessary constraints. So are the lensed images angular position, which is why a finely resolved optical exposure of the lens galaxy is required, typically obtained through space telescopes or ground telescopes with adaptive optics. To constrain the choice of a lens mass profile and subsequent gravitational potential, a few additional information need to be considered. The light profile of the lens galaxy, seen through the high-resolution image is one of them. The relative flux ratio of the lensed images as well, provided they can be corrected of the extra magnification induced by microlensing. The presence of an Einstein ring informs us on the total mass enclosed in its radius (see Eq. 2.6). Additionally, the velocity dispersion of the stars in the lens galaxy relates through the Virial theorem to the evolution of the total mass with the radius (Binney and Tremaine, 1987), a relation that yet varies with the mass profile assumption. Carefully combined, these constraints help to break the mass-sheet transformation degeneracy. Note that an efficient way to break the source-position transformation quasi degeneracy is yet to be found.

The existence of close perturbers from the main lens galaxy can also affect the time-delay distance determination. The closest ones must be explicitly included in the modeling, making use of the so-called multiplane time-delay distance equation (introduced in Schneider et al. (1992) and recently developed in McCully et al. (2014); Schneider (2014)). On the contrary to the single perturber case modeled through Eq. 4.1, the multiplane equation does depend on the underlying cosmological model chosen. This can be intuitively understood the following way: in the case of a single deflector, computing the lensing effects requires to know *where* the light rays emitted by the source cross the lens plane, which corresponds to the *observables* $\boldsymbol{\theta}_i$, also called *impact parameters*. The cosmological information is folded inside the angular diameter distances D_d, D_s and D_ds (see Sec. 4.2.4). These distances are in turn folded into the time-delay distance $D_{\Delta t}$, whose value can be completely determined by the observables and Eq. 4.1. In the case of multiple deflectors, only the impact parameters on the deflector closest to the observer are observable. The impact parameters on the deflectors located further away from the observer must be computed backwards, building on the observables measured on the deflector plane the closest to the observer. This requires to explicitly take into account the angular distances between the deflector planes (see for example Eqs. 16 and 17 of (McCully et al., 2014)), which depend on the underlying choice of cosmological model. In conclusion, if the multiplane time-delay distance equation is used in the modeling, it is important to assess that the choice of cosmological model affects only marginally the output $D_{\Delta t}$ likelihood. If not, a different likelihood must be produced for each new cosmological model considered.

The work on B1608+656 (Suyu et al., 2010) and RXJ1131-1231 (Suyu et al., 2014) as well as the fourth paper of the recent H0LiCOW series (Wong et al., 2017) on HE0435-1223 all include these recent developments, making use of the `GLEE` modeling code (Suyu and Halkola, 2010; Suyu et al., 2012) specifically developed for that purpose. For the modeling of HE0435-1223, HST imaging of the lens galaxy (see Fig. 2.7) was used to precisely measure the lensed images positions and Einstein ring radius. The ground-based Low Resolution Imaging Spectrometer (LRIS), installed on one of the Keck telescopes in Hawaii allowed to measure the velocity dispersion of the stars in the lens galaxy. Finally, the additional perturbers, namely the G1 galaxy from Fig. 4.2 has been explicitly included in the modeling through the use of the multiplane time-delay distance equation, and it was shown that the choice of a prior cosmological model had only a minor impact (<1%) on the resulting time-delay distance likelihoods. Yet, despite all these constraints there are still many parameters of the modeling that needed to be explored; for example, what kind of mass profile families to consider for the deflectors, how many external perturbers should be explicitly included in the modeling, which weight should be given to the various constraints during the modeling, etc... Figure 4.3 presents a large range of possible time-delay distance distributions $D_{\Delta t}$, depending on the various assumptions made prior to the modeling. All these possible scenario, once equivalently weighted and averaged yield the final time-delay distance distribution of HE0435-1223, to be used to infer a value for H_0. The equivalent weighting of all the possible models reflects our lack of prior information on which model should be privileged. In such a case, the most conservative assumption is indeed to consider each option as similarly plausible, and marginalize over the assumptions.

4.2.3 Line-of-sight contribution

The time-delay distance D_{Δ_t} determined through the lens modeling must also be corrected from the combined effect of all the galaxies in the line-of-sight that were not yet explicitly considered in the modeling. At first approximation, all that remaining mass can either slightly focuses or defocuses the light rays coming from the source, thus biasing the observables. That effect was rarely and poorly studied in the early decades of time-delay cosmography, until Keeton and Zabludoff (2004) raised awareness about the potential impact it could have. Fassnacht et al. (2006); Momcheva et al. (2006) demonstrated it using real-life examples, quantifying that the bias could be of 5% or greater and Hilbert et al. (2007) investigated the effect on simulated line-of-sights from the Millenium Simulation (Springel et al., 2005).

The contribution of the galaxies along the line-of-sight is folded into an *external convergence*

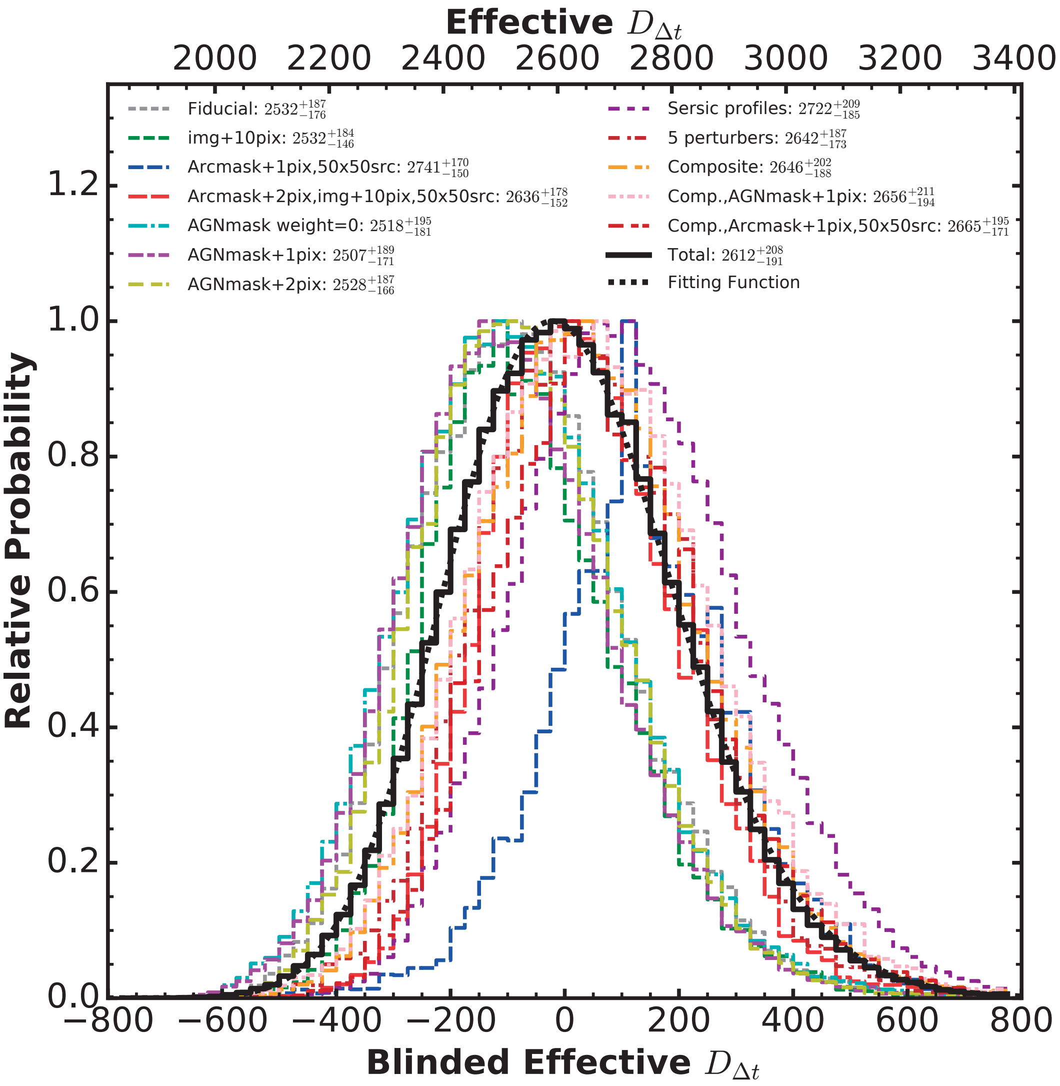


Figure 4.3: Posterior distributions of the time-delay distance $D_{\Delta t}$ computed for 12 different set of plausible modeling assumptions. The colored lines represent each individual time-delay distance (see Wong et al. (2017) for details), whereas the thick black line labeled "Total" represents the average of all models and the dotted black line proposes an analytical fit of the average distribution by a skewed log-normal function. The x-axis is blinded, meaning that each distribution has been shifted by the mean of the average distribution in order to prevent the authors of being subject to a confirmation bias during the analysis (see Sec. 4.2.1). Adapted from Wong et al. (2017).

term κ_{ext}, that directly affects time-delay distance measurement:

$$D_{\Delta_t} = \frac{D^{model}_{\Delta_t}}{1-\kappa_{ext}}, \tag{4.2}$$

where $D_{\Delta_t}^{model}$ is the time-delay distance computed through the lens modeling as presented in Sec. 4.2.2. The external convergence can be either positive or negative, depending on the additional mass distributed along the line-of-sight focuses or defocuses the light rays with respect to the case of an isolated lens galaxy.

A quantitative measure of κ_{ext} can be achieved by comparing the convergence in the target's field of view with similar field of views in numerical simulations. This is what has been achieved in Suyu et al. (2010) for the lens B1608+656, that make use of a comparison between HST observations from Fassnacht et al. (2011) and ray-tracing simulations from Hilbert et al. (2009). The approach was reproduced successfully in Suyu et al. (2013) for the analysis of RXJ1131-1231. In parallel to these results, refinements of the technique have been suggested: Greene et al. (2013) proposed a *weighted galaxy number counts* technique where the impact of galaxies is weighted by their mass, angular distance to the lens galaxy, redshift, etc... for an increased precision in determining κ_{ext}. Collett et al. (2013) proposed to go further by including a knowledge of the stellar mass versus halo mass ratio of the individual galaxies contributing to κ_{ext}. Another independent way to assess the line-of-sight contribution is to look at the weak-lensing signal in the field of the lens (see Nakajima et al., 2009; Fadely et al., 2010). That idea was developed further in McCully et al. (2014, 2017) and could provide an alternative and complementary way with *a priori* a similar precision on the measurement of κ_{ext}.

In the most recent analysis of HE0435-1223, an improved version of the weighted galaxy number count from Greene et al. (2013) was applied. It made use of wide-field, deep image of the lens galaxy surroundings to precisely and systematically identify the galaxies in the field and measure their redshift. Such a wealth of data has been acquired through multiple telescopes all around the world: namely we used a) spectroscopic data from i) the VLT in Paranal, Chile, ii) Gemini-South in Cerro Pachón, Chile, iii) the Keck telescope in Mauna Kea, Hawaii and iv) Magellan telescope in Las Campanas, Chile and b) photometric data from i) CFHT, ii) the Subaru Telescope and iii) Gemini-North, all three in Mauna Kea, Hawaii, as well as iv) the Spitzer Space Telescope. The details of the data acquisition and reduction procedures are presented in (Sluse et al., 2016; Rusu et al., 2017). In short, the idea is first to identify as many galaxies as possible in the field of view, distinguishing them from stars that have absolutely no visible impact on the convergence. This first step is done by comparing the obtained spectra for all the candidates with templates spectra of stars, galaxies, etc... This comparison also permits to determine the redshift of the galaxies with great precision, which is also a mandatory requirement for the lens modeling when the nearby perturbers are explicitly included (see Sec. 4.2.2). Figure 4.4 presents a wide view on the environment of HE0435-1223, that combines direct imaging and spectroscopy; the distinction between stars and galaxies, as well as the redshift of the galaxies are over-plotted.

With that information in hand, we can proceed to the weighted galaxy number counts and the determination of the external convergence. It requires first to compare the observed line-of-sight to other *control field* line-of-sights, in order to quantify if the former is over-dense or

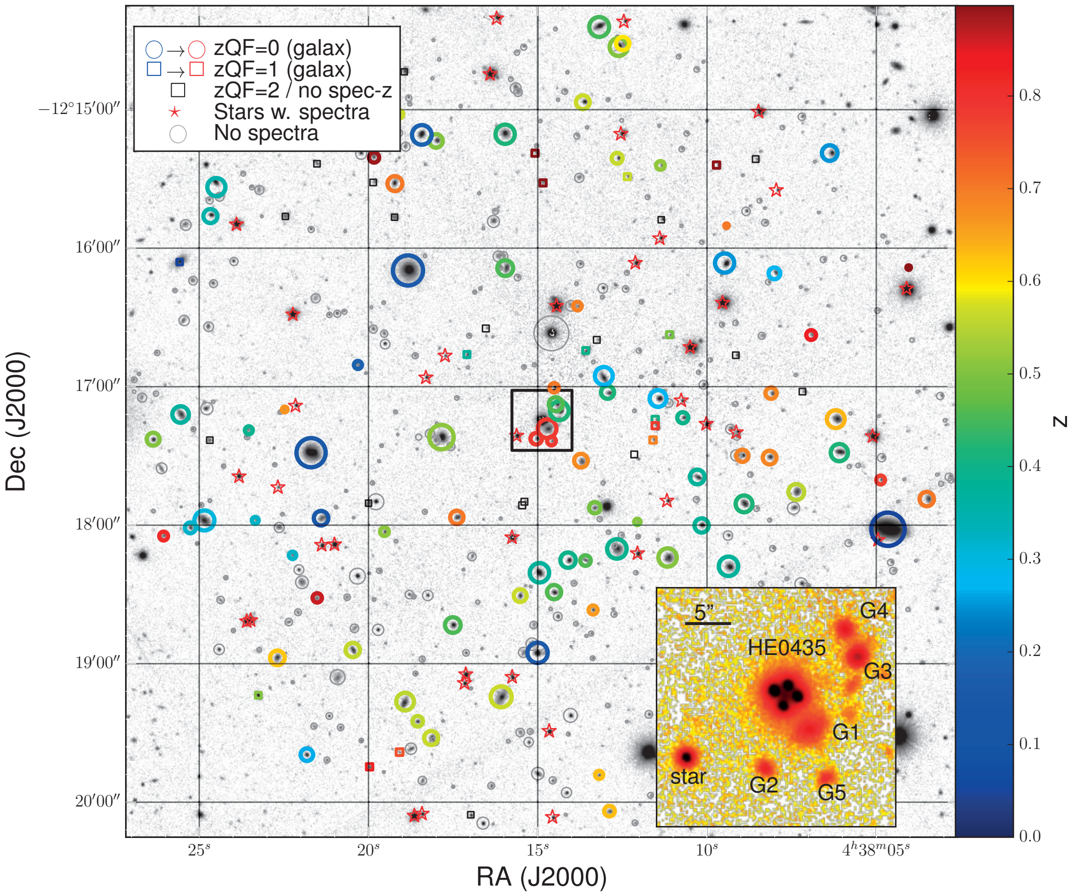


Figure 4.4: 3x3 arcminutes field of view around HE0435-1223. The over-plotted symbols indicate the nature of the objects obtained through spectroscopy: red star symbols are for stars, colored circle for galaxies. The size of the circles relates to the apparent magnitude of the galaxies, whereas the color of the circles relates to their measured redshift. Colored squares indicate galaxies with a less precise redshift measurement, where black squares show galaxies where no spectroscopic redshift measurement was successfully achieved. Finally, the gray circles indicate objects without spectra (i.e. too faint to detect a signal through the spectrographs). The bottom-right insert zooms on the lens galaxy, and corresponds to the black square at the center of the main panel, which also roughly corresponds to the field-of-view presented in Fig. 4.2. Adapted from Sluse et al. (2016).

under-dense. Then, line-of-sights of similar properties are taken from numerical simulations in order to assess the robustness of the measurement performed on real data. The tricky part of this step is to know how to properly weight the galaxies with respect to each other. There is surely more than one correct answer, which in some aspects reminds us of the choice of priors for the lens modeling detailed in Sec. 4.2.2. In that regard, what is done in Rusu et al. (2017) is to weight the galaxies according to a large range of criteria, measure in each case what is the resulting κ_{ext} and check with line-of-sights from numerical simulations if the result can be trusted. The resulting κ_{ext} and associated robustness for each choice of weights are then carefully combined in order to obtain a single probability distribution for κ_{ext} from numerical simulations. That result is finally folded into Eq. 4.2 to determine the time-delay distance probability distribution, corrected from the contribution of the galaxies along the line-of-sight. In the case of HE0435-1223, Rusu et al. (2017) find that the line-of-sight, although slightly over-dense is consistent with a $\kappa_{ext} = 0$. The eighth H0LiCOW paper (Tihhonova et al., in prep.) also tries to determine the external convergence using the weak-lensing signal from the photometric observations of the field-of-view. Its preliminary results are consistent with a null external convergence.

4.2.4 H_0 from a single lens

The time-delay distance D_{Δ_t} corrected from the external convergence κ_{ext} captures, by definition, all the cosmological information we are interested in. Let us recall its definition:

$$D_{\Delta_t} = (1+z_d)\frac{D_d D_s}{D_{ds}}, \tag{4.3}$$

where z_d is the redshift of the deflector (the lens), and D_d, D_s and D_{ds} are the angular diameter distances between the observer and the lens, the observer and the source and the lens and the source, respectively. In a ΛCDM Universe, the angular diameter distance D_{12} between two objects at redshifts z_1 and z_2 (where $z_1 < z_2$) writes as:

$$D_{12} = \frac{1}{1+z_2} f_K(x_{12}), \tag{4.4}$$

with

$$f_K(x) = \begin{cases} 1/\sqrt{K}\sin\left(\sqrt{K}\,x\right) & K > 0 \\ x & K = 0 \\ 1/\sqrt{-K}\sinh\left(\sqrt{-K}\,x\right) & K < 0 \end{cases}, \tag{4.5}$$

$$K \approx \left(\frac{H_0}{c}\right)^2 (\Omega_m + \Omega_\Lambda - 1), \tag{4.6}$$

and

$$x_{12} = \frac{c}{H_0} \int_{z_1}^{z_2} \frac{dz}{\sqrt{(1-\Omega_m-\Omega_\Lambda)(1+z)^2+\Omega_m(1+z)^3+\Omega_\Lambda}} \tag{4.7}$$

where Ω_m is the present (i.e. at $z = 0$) density of matter of the Universe normalized to its total density, and Ω_Λ is the normalized energy density associated to the vacuum - the cosmological constant -, or associated to the dark energy depending on the cosmological model considered[1]. x_{12} is called the *comoving distance* between the two objects at redshifts z_1 and z_2 and is defined such that it does *not* vary over time due to the Universe's expansion. The K factor encompasses the energy density associated to the curvature of the Universe that can be written as $\Omega_k = (1-\Omega_m-\Omega_\Lambda)$. Similarly to a two-dimensional piece of paper that can be bent in a three-dimensional space so that it closes on itself for example, a three-dimensional universe can be similarly curved. We speak about a flat, open or closed Universe for a value of Ω_k (or K) that is null, negative (positive) or positive (negative), respectively.

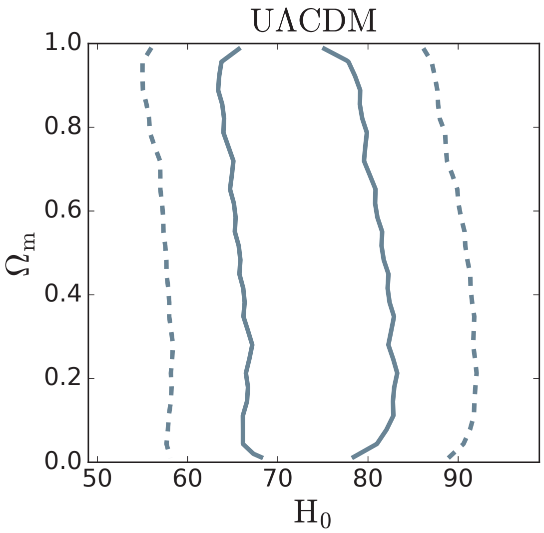
Coming back to the time-delay distance D_{Δ_t}, one can see directly from the equations above that the ratio of angular diameter distances depends primarily on H_0^{-1}. There is also a weak dependency on the square root of the Universe components density (Ω_m, Ω_Λ and Ω_k)[2]. Then, for a given D_{Δ_t} probability distribution and a choice of cosmological model, one can compute a joint probability distribution for the cosmological parameters. In a Bayesian framework, the simplest way to do it follows a grid-based approach that can be summarized as follows: chose a range of wide uniform priors for the cosmological parameters $P = \{H_0, \Omega_m, \Omega_\Lambda, \Omega_k\}$, create an uniformly spaced grid from the priors, and at each knot of the grid $P_i = \{H_{0_i}, \Omega_{m_i}, \Omega_{\Lambda_i}, \Omega_{k_i}\}$ (what we call a *sample*) compute the corresponding time-delay distance $D_{\Delta_{t,i}}$ using Eq. 4.1. In parallel to this, fit the time-delay distance distribution obtained from the lens modeling by any analytic function - typically, Suyu et al. (2010, 2014); Wong et al. (2017) proposed a skewed log-normal distribution. Then, estimate each sample $D_{\Delta_{t,i}}$ likeliness with the analytical probability function computed from the lens modeling. The resulting likeliness gives the *weight* w_i of the sample P_i. Finally, to construct a probability distribution for each cosmological parameter considered in P, simply weight each sample parameter value with the weight previously computed. In other words, the *marginalized probability distribution* corresponds to the weighted histogram of the P_i parameters' value.

The procedure described above is a specific case of the *importance sampling* technique as presented in Lewis and Bridle (2002). It slightly differs from the widely used *Markov Chain Monte-Carlo* approach (or MCMC) that does not start from a predetermined distribution of samples but generate each new sample according to the initial priors as well as the previously generated sample. A thorough comparison with the well-known MCMC sampler codes

---

[1] In such a case, we also refer to it as Ω_{de}, for dark energy.

[2] This does not hold completely true if one computes the angular diameter distances in another cosmological model than ΛCDM. In such cases, the expression of the comoving distance x_{12} has to be adapted.



cosmology: Flat ΛCDM

priors: H_0 uniform in [0,150]
Ω_Λ uniform in [0,1]
$\Omega_m = 1 - \Omega_\Lambda$

results: $H_0 = 73.1^{+5.7}_{-6.0}$
$\Omega_\Lambda = 0.51^{+0.34}_{-0.34}$

Figure 4.5: The left panel presents the posterior distribution of H_0 and Ω_m in a flat ΛCDM cosmology with uniform priors, determined from the time-delay distance distribution of HE0435-1223 presented in Fig.4.3. The solid and dashed lines represent the 68% and 95% confidence regions of the posterior distribution, respectively. The right panel presents the prior constraints of the cosmological model and marginalized H_0 and Ω_Λ values. The quoted errors correspond to the 16^{th} and 84^{th} percentiles of the marginalized distribution. Adapted from Wong et al. (2017).

`CosmoMC` (Lewis and Bridle, 2002) and `CosmoSIS` (Zuntz et al., 2015)[3] showed that the resulting probability distribution were, if not perfectly similar, extremely close to each other (G. Addison, private communication). The reason we use importance sampling instead of MCMC is that it allows a fast and easy combination of time-delay cosmography results with other data sets. Namely, CMB experiment results come in the form of lists of samples (an *MCMC chain*), each with its associated weight that allows to draw probability distribution by weighting each sample's parameters' value. Importance sampling allows us to combine two or more data sets simply by multiplying the corresponding samples' weights instead of generating a new MCMC chain.

Figure 4.5 presents the posterior distribution in the H_0 versus Ω_m plane, drawn from the HE0435-1223 constraints computed in a flat (i.e. $\Omega_k = 0$) ΛCDM Universe. It illustrates well the weak dependency of the time-delay distance to the cosmological parameters other than H_0: here, Ω_m -and consequently Ω_Λ- is very loosely constrained, meaning that the time-delay distance is not really sensitive to it. The Hubble constant, however, is much better constrained with a value of H_0 =$73.1^{+5.7}_{-6.0}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, i.e. at 8% precision. It may seem that no much progresses in terms of precision have been achieved since the previous estimates from various other lenses presented in Sec. 4.1.2, but that would be forgetting that this most

---

[3] In which the H0LiCOW results have been implemented, see https://bitbucket.org/joezuntz/cosmosis-standard-library.

recent measurement includes all the the possible source of known unknowns so far, with the sole exception of the source-plane transformation. In that regard, keeping up with the precision of the previous measurements is already a great success that augurs well for the future analysis. The fourth paper of the H0LiCOW series (Wong et al., 2017) that details the full modeling and cosmological constraints from HE0435-1223 is reproduced below.

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **465,** 4895–4913 (2017)
Advance Access publication 2016 November 29

doi:10.1093/mnras/stw3077

H0LiCOW – IV. Lens mass model of HE 0435−1223 and blind measurement of its time-delay distance for cosmology

Kenneth C. Wong,[1,2⋆†] Sherry H. Suyu,[3,2,4] Matthew W. Auger,[5] Vivien Bonvin,[6] Frederic Courbin,[6] Christopher D. Fassnacht,[7] Aleksi Halkola, Cristian E. Rusu,[7] Dominique Sluse,[8] Alessandro Sonnenfeld,[9] Tommaso Treu,[10] Thomas E. Collett,[11] Stefan Hilbert,[12,13] Leon V. E. Koopmans,[14] Philip J. Marshall[15] and Nicholas Rumbaugh[7]

[1] *National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan*
[2] *Institute of Astronomy and Astrophysics, Academia Sinica (ASIAA), PO Box 23-141, Taipei 10617, Taiwan*
[3] *Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, D-85748 Garching, Germany*
[4] *Physik-Department, Technische Universität München, James-Franck-Straße 1, D-85748 Garching, Germany*
[5] *Institute of Astronomy, University of Cambridge, Madingley Rd, Cambridge CB3 0HA, UK*
[6] *Laboratoire d'Astrophysique, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, CH-1290 Versoix, Switzerland*
[7] *Department of Physics, University of California Davis, 1 Shields Avenue, Davis, CA 95616, USA*
[8] *STAR Institute, Quartier Agora – Allée du six Août, 19c B-4000 Liège, Belgium*
[9] *Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU, WPI), University of Tokyo, Chiba 277-8583, Japan*
[10] *Department of Physics and Astronomy, University of California, Los Angeles, CA 90095-1547, USA*
[11] *Institute of Cosmology and Gravitation, University of Portsmouth, Burnaby Rd, Portsmouth PO1 3FX, UK*
[12] *Exzellenzcluster Universe, Boltzmannstr. 2, D-85748 Garching, Germany*
[13] *Ludwig-Maximilians-Universität, Universitäts-Sternwarte, Scheinerstr. 1, D-81679 München, Germany*
[14] *Kapteyn Astronomical Institute, University of Groningen, PO Box 800, NL-9700 AV Groningen, the Netherlands*
[15] *Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 452 Lomita Mall, Stanford, CA 94035, USA*

Accepted 2016 November 24. Received 2016 November 21; in original form 2016 June 30

ABSTRACT

Strong gravitational lenses with measured time delays between the multiple images allow a direct measurement of the time-delay distance to the lens, and thus a measure of cosmological parameters, particularly the Hubble constant, H_0. We present a blind lens model analysis of the quadruply imaged quasar lens HE 0435−1223 using deep *Hubble Space Telescope* imaging, updated time-delay measurements from the COSmological MOnitoring of GRAvItational Lenses (COSMOGRAIL), a measurement of the velocity dispersion of the lens galaxy based on Keck data, and a characterization of the mass distribution along the line of sight. HE 0435−1223 is the third lens analysed as a part of the H_0 Lenses in COSMOGRAIL's Wellspring (H0LiCOW) project. We account for various sources of systematic uncertainty, including the detailed treatment of nearby perturbers, the parametrization of the galaxy light and mass profile, and the regions used for lens modelling. We constrain the effective time-delay distance to be $D_{\Delta t} = 2612^{+208}_{-191}$ Mpc, a precision of 7.6 per cent. From HE 0435−1223 alone, we infer a Hubble constant of $H_0 = 73.1^{+5.7}_{-6.0}$ km s^{-1} Mpc^{-1} assuming a flat ΛCDM cosmology. The cosmographic inference based on the three lenses analysed by H0LiCOW to date is presented in a companion paper (H0LiCOW Paper V).

Key words: gravitational lensing: strong – cosmological parameters – distance scale.

1 INTRODUCTION

The flat Λ cold dark matter (ΛCDM) cosmological model is the concordance model of our Universe today. It is consistent with a variety of independent experiments, including an analysis of the cosmic microwave background (CMB) by the *Planck* mission (Planck

---

⋆E-mail: kcwong19@gmail.com
† EACOA Fellow.

Published by Oxford University Press on behalf of the Royal Astronomical Society

Collaboration XIII 2016). The *Planck* results provide the most precise cosmological parameter constraints to date, under the assumption of spatial flatness. However, there is no physical reason to assume flatness, and if the flatness assumption is relaxed, there are strong degeneracies among the cosmological parameters inferred from CMB data, particularly with the Hubble constant, H_0 (e.g. Freedman et al. 2012; Riess et al. 2016). Therefore, an independent determination of H_0 is crucial for understanding the nature of the Universe (e.g. Hu 2005; Suyu et al. 2012a; Weinberg et al. 2013).

The idea of using gravitational lens time delays to measure the Hubble constant dates back to Refsdal (1964). In practice, gravitational lens time delays provide a one-step method to determine the distance and hence the Hubble constant (e.g. Vanderriest et al. 1989; Keeton & Kochanek 1997; Schechter et al. 1997; Kochanek 2003; Koopmans et al. 2003; Saha et al. 2006; Oguri 2007; Fadely et al. 2010; Suyu et al. 2010, 2013; Sereno & Paraficz 2014; Rathna Kumar, Stalin & Prabhu 2015; Birrer, Amara & Refregier 2016; Chen et al. 2016). This method is independent of the cosmic distance ladder (e.g. Riess et al. 2011; Freedman et al. 2012) and serves as a key test of possible systematic effects in individual H_0 probes. This method rests on the fact that light rays emitted from the source at the same instant will take different paths through space–time at each of the image positions. These paths have different lengths and traverse different gravitational potentials before reaching the observer, leading to an offset in arrival times. If the source exhibits variations in its flux, the delays can be measured by monitoring the lensed images. The measured time delays can be used to calculate the time-delay distance, a combination of angular diameter distances among the observer, lens and source. The time-delay distance is primarily sensitive to H_0, with weaker dependence on other cosmological parameters (e.g. Coe & Moustakas 2009; Treu & Marshall 2016).

However, a precise and accurate determination of H_0 through this method requires a variety of observational data. A dedicated long-term monitoring campaign is necessary to obtain accurate time delays, as the uncertainty in H_0 is directly related to the relative uncertainty in the measured time delays. Deep, high-resolution imaging is required to accurately model the lens using the extended source images, which is needed to break degeneracies between the mass profile and the underlying cosmology (e.g. Kochanek 2002; Warren & Dye 2003). In order to reduce the effects of the mass sheet degeneracy (e.g. Falco, Gorenstein & Shapiro 1985; Gorenstein, Shapiro & Falco 1988; Saha 2000; Schneider & Sluse 2013; Xu et al. 2016), a measurement of the lens galaxy's velocity dispersion (e.g. Koopmans et al. 2003; Koopmans 2004) and an estimate of the external convergence, $\kappa_{\rm ext}$, along the line of sight (LOS) is needed. $\kappa_{\rm ext}$ can also bias the lens model parameters if unaccounted for (e.g. Collett et al. 2013; Greene et al. 2013; McCully et al. 2014, 2016).

In an effort to provide an accurate independent estimate of H_0 using time-delay lenses, we use a number of new data sets as part of our project, H_0 Lenses in COSMOGRAIL's Wellspring (H0LiCOW), to model five lensed quasars. These data sets include high-resolution imaging with the *Hubble Space Telescope* (*HST*), precise time-delay measurements from the COSmological MOnitoring of GRAvItational Lenses (COSMOGRAIL; Courbin et al. 2005; Eigenbrod et al. 2005; Bonvin et al. 2016b) project and from Very Large Array monitoring (Fassnacht et al. 2002), a photometric and spectroscopic survey to characterize the LOS mass distribution to estimate $\kappa_{\rm ext}$ in these systems and stellar velocity dispersion measurements of the strong lens galaxies. With five separate lenses, we plan to account for systematic uncertainties and obtain a robust constraint on H_0 to <3.5 per cent precision.

In this paper, we present the results of a detailed lens modelling analysis of the gravitational lens HE 0435−1223 using new high-resolution imaging data from *HST*. HE 0435−1223 is the third H0LiCOW system analysed in this manner, following B1608+656 (Suyu et al. 2010) and RXJ1131−1231 (Suyu et al. 2013, 2014). This paper is the fourth in a series of papers detailing our analysis of HE 0435−1223. The other papers include an overview of the H0LiCOW project (Suyu et al. 2016, hereafter H0LiCOW Paper I), a spectroscopic survey of the HE 0435−1223 field and a characterization of the groups along the LOS (Sluse et al. 2016, hereafter H0LiCOW Paper II), a photometric survey of the HE 0435−1223 field and an estimate of $\kappa_{\rm ext}$ due to the external LOS structure (Rusu et al. 2016, hereafter H0LiCOW Paper III), and a presentation of our latest time-delay measurements for HE 0435−1223 and the cosmological inference from our combined analysis of HE 0435−1223, B1608+656and RXJ1131−1231 (Bonvin et al. 2016a, hereafter H0LiCOW Paper V).

This paper is organized as follows. We provide a brief overview of using time-delay lenses for cosmography in Section 2. In Section 3, we describe the observational data used in our analysis. We describe our lens modelling procedure in Section 4. The time-delay distance results and their implications for cosmology are presented in Section 5. We summarize our main conclusions in Section 6. Throughout this paper, all magnitudes given are on the AB system.

2 TIME-DELAY COSMOGRAPHY

2.1 Time-delay distance

When a source is gravitationally lensed, the light travel time from the source to the observer depends on both the path length of the light rays and the gravitational potential of the lens through which the rays pass. For a single lens plane, the excess time delay of an image at an angular position $\boldsymbol{\theta} = (\theta_1, \theta_2)$ with corresponding source position $\boldsymbol{\beta} = (\beta_1, \beta_2)$ relative to the case of no lensing is

$$t(\boldsymbol{\theta}, \boldsymbol{\beta}) = \frac{D_{\Delta t}}{c} \left[\frac{(\boldsymbol{\theta} - \boldsymbol{\beta})^2}{2} - \psi(\boldsymbol{\theta}) \right], \qquad (1)$$

where $D_{\Delta t}$ is the time-delay distance and $\psi(\boldsymbol{\theta})$ is the lens potential. The time-delay distance (Refsdal 1964; Schneider, Ehlers & Falco 1992; Suyu et al. 2010) is defined[1] as

$$D_{\Delta t} \equiv (1 + z_{\rm d}) \frac{D_{\rm d} D_{\rm s}}{D_{\rm ds}}, \qquad (2)$$

where $z_{\rm d}$ is the lens redshift, $D_{\rm d}$ is the angular diameter distance to the lens, $D_{\rm s}$ is the angular diameter distance to the source, and $D_{\rm ds}$ is the angular diameter distance between the lens and the source. Since $D_{\Delta t}$ has units of distance, it is inversely proportional to H_0.

For lens systems with multiple deflectors at distinct redshifts, the observed time delays depend on various combinations of the angular diameter distances measured between us, the multiple deflectors and the source, and the observed time delays are no longer proportional to a single time-delay distance. The observed image positions depend on the multiplane lens equation (e.g. Blandford & Narayan 1986; Kovner 1987; Schneider et al. 1992; Petters, Levine & Wambsganss 2001; Collett & Auger 2014; McCully et al. 2014). However, for a system where the lensing is dominated by a single

---

[1] For historical reasons, the time-delay distance is written in terms of angular diameter distances. A more natural definition is $D_{\Delta t} \equiv \hat{D}_{\rm d} \hat{D}_{\rm s} / \hat{D}_{\rm ds}$ where $\hat{D}$ are the proper distances that the photons have travelled.

plane, the observed time delays are primarily sensitive to the time-delay distance defined in equation (2), with the deflector redshift as that of the primary strong lens plane. We show in Section 4.9 that this approximation is valid for HE 0435−1223 and our results can thus be interpreted as a constraint on $D_{\Delta t}(z_{\rm d}, z_{\rm s})$, which we refer to as the effective time-delay distance measured by this system. Hereafter, $D_{\Delta t}$ refers to this effective time-delay distance unless otherwise indicated.

For variable sources such as active galactic nuclei (AGN), it is possible to monitor the fluxes of the lensed images at positions $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$ and measure the time delay, $\Delta t_{ij} \equiv t(\boldsymbol{\theta}_i, \boldsymbol{\beta}) - t(\boldsymbol{\theta}_j, \boldsymbol{\beta})$, between them (e.g. Vanderriest et al. 1989; Schechter et al. 1997; Fassnacht et al. 1999, 2002; Kochanek et al. 2006; Courbin et al. 2011). The lens potentials at the two image positions, $\psi(\boldsymbol{\theta}_i)$ and $\psi(\boldsymbol{\theta}_j)$, as well as the source position, $\boldsymbol{\beta}$, can be determined from a mass model of the system. Therefore, lenses with measured time delays and accurate lens models can be used to constrain $D_{\Delta t}$.

A complicating factor in using time-delay lenses for cosmography is the fact that all mass along the LOS contributes to the lens potential that the light rays pass through. These external perturbers not only affect the lens model of the system, but also lead to additional focusing and defocusing of the light rays, which in turn affects the measured time delays (e.g. Seljak 1994). If unaccounted for, these external perturbers can lead to biased inferences of $D_{\Delta t}$. If effects of LOS perturbers are small, they can be approximated by an external convergence term in the lens plane, $\kappa_{\rm ext}$ (neglecting the $1 - \beta$ terms that enter into a more accurate prescription; Keeton 2003; McCully et al. 2014). The true $D_{\Delta t}$ is related to the $D_{\Delta t}^{\rm model}$ inferred from a mass model that does not account for $\kappa_{\rm ext}$ by

$$D_{\Delta t} = \frac{D_{\Delta t}^{\rm model}}{1 - \kappa_{\rm ext}}. \tag{3}$$

$\kappa_{\rm ext}$ cannot be constrained from the lens model due to the mass sheet degeneracy (e.g. Falco et al. 1985; Gorenstein et al. 1988; Saha 2000), in which the addition of a uniform mass sheet and a rescaling of the source plane coordinates can affect the product of the time delays and H_0 but leave other observables unchanged.

The above degeneracy caused by $\kappa_{\rm ext}$ can be broken or substantially mitigated by estimating the mass distribution along the LOS (e.g. Fassnacht et al. 2006; Momcheva et al. 2006, 2015; Williams et al. 2006; Wong et al. 2011). However, for perturbers that are very massive or projected very close to the lens, they may need to be included explicitly in the mass model, as their higher order effects need to be properly accounted for (McCully et al. 2016). On the other hand, the lens profile is also degenerate with the time-delay distance in that the radial profile slope is tightly correlated with the time-delay distance (e.g. Kochanek 2002; Wucknitz 2002; Suyu 2012). The profile degeneracy affects models that share the same form of mass density profile (e.g. a power-law density profile), as well as models with different density profiles (described analytically or not). Furthermore, the profile degeneracy can mimic the effects of the mass sheet degeneracy since different profiles can exactly or approximately be mass sheet transformations of one form or another (e.g. Schneider & Sluse 2013, 2014; Unruh, Schneider & Sluse 2016). With reasonable assumptions about the lens galaxy's mass profile, these degeneracies can be reduced by augmenting the lensing data with stellar kinematics measurements of the lens galaxy (e.g. Treu & Koopmans 2002; Koopmans et al. 2003; Auger et al. 2010; Suyu et al. 2014). Including the velocity dispersion in the modelling helps to constrain any internal uniform mass component from a local galaxy group that the dynamics is sensitive to (Koopmans 2004).

2.2 Joint inference

Our inference of $D_{\Delta t}$ follows that of Suyu et al. (2013), but with some important modifications. Our observational data sets are denoted by $\boldsymbol{d}_{\bf HST}$ for the *HST* imaging data, $\boldsymbol{\Delta t}$ for the time delays, σ for the velocity dispersion of the lens galaxy and $\boldsymbol{d}_{\bf LOS}$ for the properties of the LOS mass distribution determined from our photometric and spectroscopic data. We want to obtain the posterior probability distribution function (PDF) of the model parameters $\boldsymbol{\nu}$ given the data, $P(\boldsymbol{\xi}|\boldsymbol{d}_{\bf HST}, \boldsymbol{\Delta t}, \boldsymbol{\sigma}, \boldsymbol{d}_{\bf LOS}, \boldsymbol{A})$. The vector $\boldsymbol{\xi}$ includes the lens model parameters $\boldsymbol{\nu}$, the cosmological parameters $\boldsymbol{\pi}$ (Section 4.9) and nuisance parameters representing the external convergence ($\kappa_{\rm ext}$; Section 4.4) and anisotropy radius ($r_{\rm ani}$; Section 4.3), each of which we introduce and discuss in the sections indicated. $\boldsymbol{A}$ denotes a discrete set of assumptions we make about the form of the model, including the choices we have to make about the data modelling region, the setup of the source reconstruction grid, the treatment of the various deflector mass distributions, etc. In general, $\boldsymbol{A}$ cannot be fully captured by continuous parameters. By Bayes' theorem, we have that

$$\begin{aligned} &P(\boldsymbol{\xi}|\boldsymbol{d}_{\bf HST}, \boldsymbol{\Delta t}, \boldsymbol{\sigma}, \boldsymbol{d}_{\bf LOS}, \boldsymbol{A}) \\ &\quad \propto P(\boldsymbol{d}_{\bf HST}, \boldsymbol{\Delta t}, \boldsymbol{\sigma}, \boldsymbol{d}_{\bf LOS}|\boldsymbol{\xi}, \boldsymbol{A}) P(\boldsymbol{\xi}|\boldsymbol{A}), \end{aligned} \tag{4}$$

where $P(\boldsymbol{d}_{\bf HST}, \boldsymbol{\Delta t}, \boldsymbol{\sigma}, \boldsymbol{d}_{\bf LOS}|\boldsymbol{\xi}, \boldsymbol{A})$ is the joint likelihood function and $P(\boldsymbol{\xi}|\boldsymbol{A})$ is the prior PDF for the parameters given our assumptions. Since the data sets are independent, the likelihood can be separated,

$$\begin{aligned} P(\boldsymbol{d}_{\bf HST}, \boldsymbol{\Delta t}, \boldsymbol{\sigma}, \boldsymbol{d}_{\bf LOS}|\boldsymbol{\xi}, \boldsymbol{A}) = {} & P(\boldsymbol{d}_{\bf HST}|\boldsymbol{\xi}, \boldsymbol{A}) \\ & \times P(\boldsymbol{\Delta t}|\boldsymbol{\xi}, \boldsymbol{A}) \\ & \times P(\boldsymbol{\sigma}|\boldsymbol{\xi}, \boldsymbol{A}) \\ & \times P(\boldsymbol{d}_{\bf LOS}|\boldsymbol{\xi}, \boldsymbol{A}). \end{aligned} \tag{5}$$

We note that equation (5) assumes the approximation that the LOS can be decoupled from the lens model. We can calculate the individual likelihoods separately and combine them as in equation (5) to get the final posterior PDF for a given set of assumptions.

In Section 4.7, we lay out a range of systematics tests where we vary the content of $\boldsymbol{A}$ and repeat the inference of $\boldsymbol{\xi}$. Such a sensitivity analysis is important for checking the magnitude of various known but unmodelled systematic effects, but it leaves us with the question of how to combine the results. We note that the marginalization integral over these assumptions can be approximated as a sum as follows (denoting all four data sets by $\boldsymbol{d}$),

$$\begin{aligned} P(\boldsymbol{\xi}|\boldsymbol{d}) &= \int P(\boldsymbol{\xi}|\boldsymbol{d}, \boldsymbol{A}) P(\boldsymbol{A}|\boldsymbol{d})\, {\rm d}\boldsymbol{A} \\ &\overset{\propto}{\sim} \sum_k P(\boldsymbol{\xi}|\boldsymbol{d}, \boldsymbol{A}_k) P(\boldsymbol{d}|\boldsymbol{A}_k) \\ &\overset{\propto}{\sim} \sum_k P(\boldsymbol{\xi}|\boldsymbol{d}, \boldsymbol{A}_k), \end{aligned} \tag{6}$$

provided the following two statements are true: first, that the prior PDF over possible assumptions is uniform, and that our sampling of possible assumptions is fair. We choose reasonable variations in the systematic effects to try to achieve this. The second is that the evidence $P(\boldsymbol{d}|\boldsymbol{A}_k)$ does not change appreciably between inferences; this is likely to be true if the goodness of fit does not change, and the parameter priors and volumes are not very different. Under these assumptions, equation (6) shows that a *sum* of the posterior PDFs is

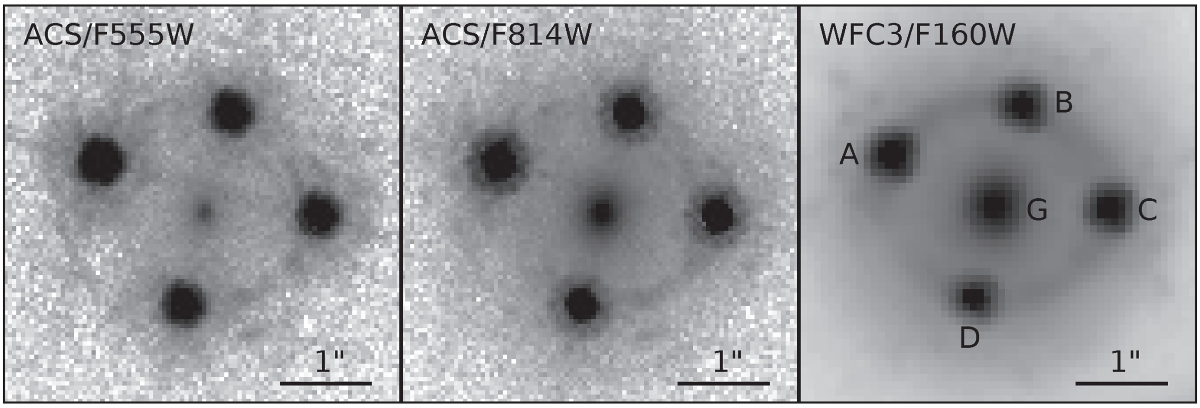


Figure 1. *HST* images of HE 0435−1223. Shown are cutouts of the lens system used for lens modelling in the ACS/*F*555*W* (left), ACS/*F*814*W* (middle) and WFC3/*F*160*W* (right) bands. The images are 4.5 arcsec on a side. The scale is indicated in the bottom right of each panel. The main lens galaxy (G) and lensed quasar images (A, B, C, D) are marked.

an approximation to the posterior PDF marginalized over the tested systematic effects.

3 DATA

HE 0435−1223 (J2000: $4^{\mathrm{h}}38^{\mathrm{m}}14\overset{\mathrm{s}}{.}9$, $-12^{\circ}17'14\overset{''}{.}4$) is a quadruply lensed quasar discovered by Wisotzki et al. (2002) as part of the Hamburg/ESO survey for bright QSOs (Wisotzki et al. 2000). The main deflector is a massive elliptical galaxy at a redshift of $z_{\mathrm{d}} = 0.4546 \pm 0.0002$ (Morgan et al. 2005), and the source redshift is $z_{\mathrm{s}} = 1.689$.[2] Our spectroscopic observations reveal that the lens is a part of a galaxy group with a velocity dispersion of $\sigma = 471 \pm 100\,\mathrm{km\,s^{-1}}$ measured from 12 member galaxies (H0LiCOW Paper II), which is independently confirmed by Wong et al. (2011) and Wilson et al. (in preparation) based on a spectroscopic study by Momcheva et al. (2006, 2015). We present the *HST* imaging used for lens modelling in Section 3.1, the time delays measured by COSMOGRAIL in Section 3.2, the spectroscopy of the lens galaxy for measuring the lens stellar velocity dispersion in Section 3.3 and ground-based imaging and spectroscopy to characterize the lens environment in Section 3.4.

3.1 *HST* imaging

We obtain deep *HST* observations of HE 0435−1223 using the Wide Field Camera 3 (WFC3) IR channel in the *F*160*W* band (Program #12889; PI: Suyu). The details of these observations are presented in H0LiCOW Paper I, which we summarize here. Using a combination of short (44 s) and long (599 s) exposures, we reconstruct the brightness distribution of both the lensed AGN and host galaxy. We reduce the images using DRIZZLEPAC.[3] The images are drizzled to a final pixel scale of 0.08 arcsec without masking the bright AGN pixels, as they are well characterized.

We also use archival observations from the Advanced Camera for Surveys (ACS) on *HST* in the *F*555*W* and *F*814*W* filters (Program #9744; PI: Kochanek). The images are reduced using MULTIDRIZZLE[4] with charge transfer inefficiency taken into account (e.g. Anderson & Bedin 2010; Massey et al. 2010). The final pixel scale of the reduced images is 0.05 arcsec.

We create cutouts of the *HST* images around the lens and define an arcmask in each band in which we perform the modelling. For the ACS bands, we use a 90 × 90 pixel cutout (4.5 arcsec on a side), and for the WFC3/*F*160*W* band, we use a 60 × 60 pixel cutout (4.8 arcsec on a side). These cutouts are shown in Fig. 1.

To generate the initial point spread function (PSF) of the exposures, we first select three stars in the field that are close to the lens galaxy in angular separation to minimize CCD distortion effects, and which have approximately the same brightness as the lensed AGN images to avoid any PSF broadening effects. We then simultaneously fit these stars with a Moffat profile plus a regularized fine-pixel array. The exposures are sky-subtracted prior to the PSF fitting. The details of this fitting procedure are described by Cantale et al. (2016a) and are based on ideas presented in Magain, Courbin & Sohy (1998). A successful application of the procedure is presented by Cantale et al. (2016b). We then use this initial PSF as the starting point for our iterative PSF correction procedure (see Section 4.1).

The weight images are constructed as follows. We take a large, relatively sparse area of the image and approximate the background noise as the normalized median absolute deviation (NMAD), defined as $\mathrm{NMAD} \equiv 1.48 \times \mathrm{median}(|p_i - \mathrm{median}(p)|)$, where p_i is the value of pixel i and $\mathrm{median}(p)$ is the median of all pixels in the selected area. We use the NMAD, which is a good approximation to the standard deviation, as it is less sensitive to outliers. We create a 'noise image' that has the same dimensions as the lens galaxy cutout with all pixels initialized to the value of the background noise. We then add Poisson noise to this noise image by taking all pixels in the lens galaxy cutout where the flux is greater than the background noise level and adding in quadrature the square root of each pixel value (normalized by its effective exposure time) to the corresponding pixel in the noise image (this is because the units of the science image are counts per second). The noise image is then squared and inverted to obtain the weight image. We note that

---

[2] We note that Sluse et al. (2012) measure an updated source redshift of $z_{\mathrm{s}} = 1.693$. We use the original value of $z_{\mathrm{s}} = 1.689$ in our analysis but verify that using this updated measurement does not impact our results.

[3] DRIZZLEPAC is a product of the Space Telescope Science Institute, which is operated by AURA for NASA.

[4] MULTIDRIZZLE is a product of the Space Telescope Science Institute, which is operated by AURA for NASA.

while the background noise for the WFC3 IR camera depends on the number of non-destructive reads, we verify that the number of reads in the region of the lensed arc is the same as for the blank sky patch used for estimating the background noise, so this procedure is valid.

When modelling with these weights, there are large residuals near the AGN image centres due to our inability to model the PSF on a grid of pixels with sufficient accuracy. This can lead to biased results as the model will be influenced by these relatively small areas rather than the large-scale features of the source, so we compensate for this by reducing the weight in these regions (e.g. Suyu 2012). We scale the weight in these regions by a power law such that a pixel originally given a noise value of p_i is rescaled to a noise value of $A \times p_i^b$. The A and b are constants that are different for each band and are chosen such that the normalized residuals in the AGN image regions are approximately consistent with the normalized residuals in the rest of the arc region.

We note that in determining the effective exposure time on a pixel-by-pixel basis, we turn off the bad pixel masking in a 3 × 3 pixel region around the brightest pixel of each of the lensed AGN images. This is done because allowing bad pixel masking results in interpolations of the image pixels that cause the four AGN images to exhibit different PSF profiles, which complicates our iterative PSF correction scheme (Section 4.1). Turning off the bad pixel mask produces more faithfully and consistently the four AGN images. Since the majority of the lens mass model constraints come from the lensing arcs away from the centres of the AGN images, we have checked that these arcs do not have bad pixels that would affect our lens mass model.

3.2 Time-delay measurements

Time-delay measurements for HE 0435−1223 were initially given in Courbin et al. (2011). Further monitoring of the system by COSMOGRAIL has since improved the time-delay accuracy and precision, completing the data from Courbin et al. (2011) with ∼1300 exposures of 6 min each for a total of 301 new observing nights ranging from 2010 September to 2016 April. The details of the data acquisition and time-delay measurements used in our analysis are presented in H0LiCOW Paper V, but we summarize the main results here.

The data treatment follows the procedure described by Tewes et al. (2013b). Each observing epoch is corrected following the standard reduction steps (bias subtraction, flat-fielding and sky correction). The PSF is estimated following the procedure described in Section 3.1. The exposures are then normalized using bright, non-saturated stars in the field of view. The photometry of the four images of HE 0435−1223 is obtained on each exposure using the Magain et al. (1998) deconvolution photometry presented in Cantale et al. (2016b). The light curves obtained with this method are presented in fig. 2 of H0LiCOW Paper V.

The measurement of the time delays between each pair of images follows the formalism introduced by Tewes, Courbin & Meylan (2013a). The common intrinsic variability of the quasar and the four independent extrinsic variability curves are fitted using free-knot splines. The curves are then shifted in time to optimize the fit. The uncertainties on the time-delay measurements are estimated using a Monte Carlo approach. A set of 1000 synthetic light curves are drawn, mimicking the light curves and the time-delay constraining power of the observed data (Tewes et al. 2013a). It is important that the synthetic data sets span a range of plausible true time delays, as this allows us to verify that the estimator accurately responds to theses input delays (i.e. does not suffer from lethargy, described in Rathna Kumar et al. 2015) and has not been involuntarily fine-tuned to recover a particular value of the time delay. Various tests on the data reduction process and curve-shifting technique have been performed successfully to ensure the reliability of the time-delay measurements. We use the time delays relative to image A: $\Delta t_{\rm AB} = -8.8 \pm 0.8$ d, $\Delta t_{\rm AC} = -1.1 \pm 0.7$ d and $\Delta t_{\rm AD} = -13.8 \pm 0.9$ d, where the uncertainties represent 1σ confidence intervals.

3.3 Stellar velocity dispersion of lens galaxy

HE 0435−1223 was observed with the Low-Resolution Imaging Spectrometer (LRIS; Oke et al. 1995) on the Keck I telescope on 2011 January 4. Six exposures of 1200 s were obtained in 0.8 arcsec seeing with the red arm of the spectrograph using the 831/8200 grating, which has a dispersion of 0.58 Å pixel^{-1} and yields an effective resolution $\sigma_{\rm res} \sim 37$ km s^{-1}. The 0.75 arcsec slit was oriented to intersect the eastern- and western-most lensed QSO images (i.e. at a position angle of 76°) and a 4-pixel (0.54 arcsec) aperture was used to extract 1D spectra from each exposure. These six spectra were then resampled to a single spectrum using spline interpolation and rejecting pixels affected by cosmic rays or other artefacts; the resulting spectrum is shown in Fig. 2. The velocity dispersion was obtained following the same procedure as in Suyu et al. (2010, 2013), resulting in an inference of $\sigma = 222$ km s^{-1} with a statistical uncertainty of 11 km s^{-1} and a systematic uncertainty of ∼10 km s^{-1} due to the templates used, the region of the spectrum that was fitted, and the order of the polynomial continuum. We therefore adopt an overall uncertainty of $\sigma_\sigma = 15$ km s^{-1}. This measurement is in agreement with a previous determination of $\sigma = 222 \pm 34$ km s^{-1} by Courbin et al. (2011) within a 1 arcsec aperture.

3.4 Lens environment: photometry and spectroscopy

To account for the effects of LOS structure, we have obtained deep multiband photometry and multi-object spectroscopy in the HE 0435−1223 field to characterize the external mass distribution. Details of the photometric observations and inference on $\kappa_{\rm ext}$ are presented in H0LiCOW Paper III, and the details of the spectroscopic data are presented in H0LiCOW Paper II, but we summarize the data here.

Our wide-field photometric data consist of ground-based *ugriJHK*$_{\rm s}$ observations, as well as 3.6, 4.5, 5.8 and 8.0 μm observations with the *Spitzer* Infrared Array Camera (IRAC; Fazio et al. 2004). We infer photometric redshifts and stellar masses using PSF-matched photometry measured with SExtractor. We use LEPHARE (Ilbert et al. 2006) to measure stellar masses for the best-fitting redshift using the spectral energy distribution (SED) templates employed by CFHTLenS (Velander et al. 2014), which assume a Chabrier (2003) initial mass function.

The wide-field spectroscopic data are taken with a combination of Keck/LRIS, the Focal Reducer/low-dispersion Spectrograph 2 (FORS2; Appenzeller et al. 1998) on the Very Large Telescope (VLT), and the Gemini Multi-Object Spectrograph (GMOS; Hook et al. 2004), and are combined with existing spectroscopic observations of this field (Momcheva et al. 2006, 2015). It is particularly important to model the most significant perturbers, as their effects may not be adequately accounted for by external shear alone (McCully et al. 2014). McCully et al. (2016) find that the most significant perturbers are those that are massive, projected close to the lens, and that are in the foreground of the lens redshift. H0LiCOW Paper II presents an estimate of the relative significance of nearby perturbers to HE 0435−1223 as quantified by their flexion shift,

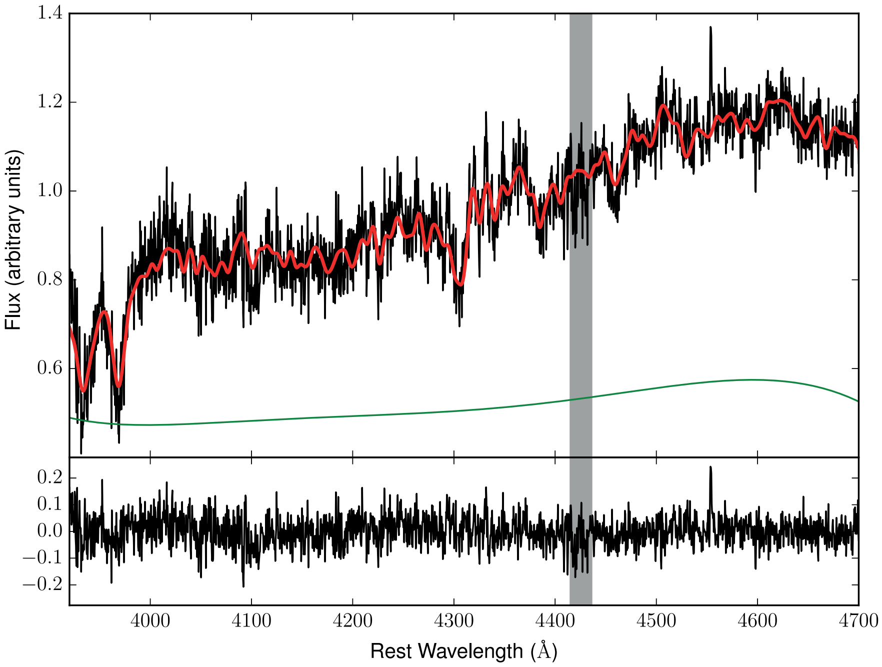


Figure 2. Top: Keck/LRIS spectrum of HE 0435−1223 with the best-fitting model overplotted in red and a polynomial continuum, which accounts for contamination from the lensed QSO images and template mismatch, shown in green. We find that $\sigma = 222 \pm 15$ km s^{-1}, including systematic uncertainties due to the templates used, the region of the spectrum that was fitted, and the order of the polynomial continuum. The grey vertical band represents a wavelength range that is excluded from the fit due to the presence of a strong Mg II absorption system. Bottom: residuals from the best fit.

$\Delta_3 x$ (McCully et al. 2016), and finds that at most, the five nearest perturbers should be accounted for explicitly, with all other perturbers having a negligible influence. Fig. 3 shows the lens and the relative positions and redshifts of these five perturbers, all brighter than $i = 22.5$ mag and projected within 12 arcsec of the lens.

4 LENS MODELLING

In this section, we describe our procedure to simultaneously model the images in the three *HST* bands and the time delays to infer the lens model parameters.

4.1 Overview

We perform our lens modelling using GLEE, a software package developed by S. H. Suyu and A. Halkola (Suyu & Halkola 2010; Suyu et al. 2012b). The lensing mass distribution is described by a parametrized profile. The extended host galaxy of the source is modelled separately on a 40 × 40 pixel grid with curvature regularization (Suyu et al. 2006). The lensed quasar images are modelled as point sources convolved with the PSF. By modelling the quasar images on the image plane independently from the extended host galaxy light distribution, we allow for variations in quasar fluxes due to microlensing, time delays and substructure. The lens galaxy light distribution is modelled using Chameleon profiles (defined as the difference of two non-singular r^{-2} elliptical profiles; Kassiola & Kovner 1993; Dutton et al. 2011), which are a good approximation to Sérsic profiles. We represent the galaxy light distribution as the sum of two Chameleon profiles with a common centroid. We use Chameleon profiles rather than Sérsic profiles because they provide a similarly good fit to the data (see Sections 4.7 and 5) and it is

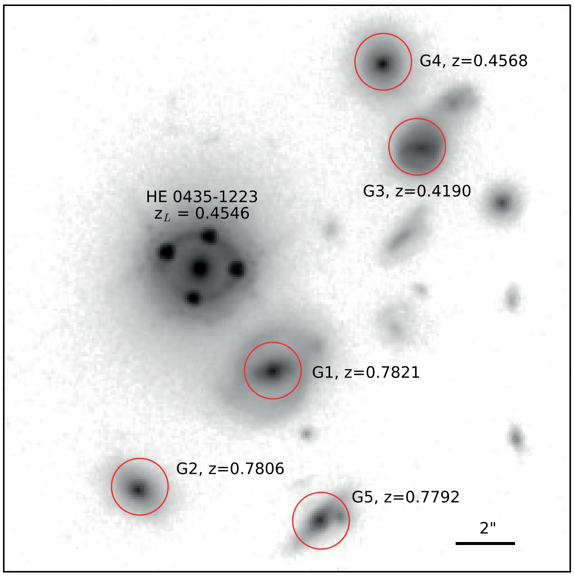


Figure 3. *HST*/WFC3 *F*160*W* image of a 20 arcsec × 20 arcsec field around HE 0435−1223. The angular scale is indicated in the bottom right corner. The five most significant nearby perturbers are marked with red circles, and the redshifts of the perturbers are indicated. G1 is included explicitly in our model, as it is the most massive and nearest in projection to HE 0435−1223. We also test the effects of including the other perturbers as one of our systematics tests.

more straightforward to link their parameters to the mass parameters in our tests of alternative mass models. Model parameters of the lens and source are constrained through Markov Chain Monte Carlo (MCMC) sampling.

Since we account for G1 at a different redshift from the main lens galaxy, we make use of the full multiplane lens equation (e.g. Blandford & Narayan 1986; Kovner 1987; Schneider et al. 1992; Petters et al. 2001; Collett & Auger 2014; McCully et al. 2014) in our modelling. We vary H_0 directly in our models and use this distribution to calculate the effective model time-delay distance $D_{\Delta t}^{\rm model}$. In calculating $D_{\Delta t}^{\rm model}$, we assume $\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$ and $w = -1$, although we show that relaxing these assumptions shifts the resulting $D_{\Delta t}^{\rm model}$ distributions by <1 per cent (Section 4.9).

4.2 Mass model

Our primary mass model for the lens galaxy is a singular power-law elliptical mass distribution (hereafter 'SPEMD'; Barkana 1998), although we also test a model consisting of a baryonic component that traces the light distribution and a separate dark matter component (hereafter the 'composite' model; see Section 4.7). We also include an external shear in the strong lens plane. Past studies have shown that a power-law model provides a good general description of typical lens galaxies at the length-scales we are interested in (e.g. Koopmans et al. 2006, 2009; Suyu et al. 2009; Auger et al. 2010; Barnabè et al. 2011; Sonnenfeld et al. 2013).

We also explicitly include the most nearby massive perturbing galaxy [G1 in Fig. 3; $z = 0.7821$, $\log(M_*/{\rm M}_\odot) = 10.9$] that is projected ~4.5 arcsec away from the lens, which is close enough that its influence may not be adequately described by external shear (H0LiCOW Paper II; see also McCully et al. 2016). G1 is modelled as a singular isothermal sphere, which is a reasonable assumption as higher order moments of its potential will have a small influence at the position of the main lens galaxy. G1 is treated using the full multiplane lens equation, as detailed by Suyu et al. (in preparation).

Our SPEMD model has the following free parameters:

(i) position (θ_1, θ_2) of the centroid (allowed to vary independently from the centroid of the light distribution);
(ii) *Einstein* radius $\theta_{\rm E}$;
(iii) minor-to-major axis ratio q and associated position angle θ_q;
(iv) 3D slope of the power-law mass distribution γ';
(v) external shear $\gamma_{\rm ext}$ and associated position angle θ_γ;[5]
(vi) *Einstein* radius of G1;
(vii) the cosmological parameter H_0.

In principle, our lens is drawn from a selection function and the choice of model priors may introduce a bias on the inferred time-delay distance. However, since the selection function is not well known and these biases are negligibly small for an analysis like ours (Collett & Cunnington 2016), we conservatively assume uniform priors on the model parameters.

To get a starting point for our model, we run a preliminary model where only the positions and time delays of the lensed quasar images are used as constraints and G1 is not included. This preliminary model is fast and easy to optimize, and we use the output parameters as the initial parameters of our primary model.

Our constraints on the primary lens model include the positions of the lensed quasar images, the measured time delays and the surface brightness of the pixels in the ACS/*F*555*W*, ACS/*F*814*W* and WFC3/*F*160*W* images that are fit simultaneously. We first model the lens system individually in each band to iteratively update the PSFs using the lensed AGN images themselves in a manner similar to Chen et al. (2016), but with the PSF corrections and source intensity reconstructed simultaneously in our case rather than separately (Suyu et al., in preparation). We then fix these 'corrected' PSFs and use them in our final models that simultaneously use the surface brightness distribution in all three bands as constraints. We do not enforce any similarity of pixel values at the same spatial position across different bands. In our MCMC sampling, we vary the light parameters of the lens galaxy and quasar images, the mass parameters of the lens galaxy, the external shear, the *Einstein* radius of G1 and H_0. The quasar positions are linked across all three bands, but the other light parameters are allowed to vary independently.

Fig. 4 shows the data and the lens model results in each of the three bands, as well as the source reconstruction. Our model reproduces the surface brightness structure of the lensed AGN and host galaxy in all three bands. There are some small residuals in the region of the lensed arc away from the AGN images. We attribute these to compact star formation regions in the host galaxy, as our model maps these features to similar locations in the source plane. We test a model where the region near these residuals are masked out and find that our $D_{\Delta t}$ inference is consistent to within our systematic uncertainties (Section 4.7).

4.3 Kinematics

We follow Suyu et al. (2010) and Sonnenfeld et al. (2012) to compute the LOS stellar velocity dispersion of the strong lens galaxy through the spherical Jeans equation (see also Treu & Koopmans 2002; Koopmans et al. 2003). For a given lens model, we obtain the 3D density profile of the lens galaxy by taking the spherical deprojection of the circularized surface mass density profile. The resulting 3D density profile assumes an analytical form for both the power-law and the composite model. The 3D distribution of tracers is obtained by applying the same procedure to the surface brightness distribution of the lens galaxy, which we model as a Hernquist (1990) profile. We also tested a Jaffe (1983) profile that has been shown to produce similar results (Suyu et al. 2010), and find that the results are affected by less than 1 per cent level. We parametrize the orbital anisotropy profile with an Osipkov–Merritt model (Osipkov 1979; Merritt 1985)

$$\frac{\sigma_\theta^2}{\sigma_r^2} = 1 - \frac{r^2}{r_{\rm ani}^2 + r^2}. \tag{7}$$

Given values of the lens mass parameters in Section 4.2, the external convergence $\kappa_{\rm ext}$ in Section 4.4 and the anisotropy radius $r_{\rm ani}$, we then calculate the LOS velocity dispersion profile by numerically integrating the solutions of the spherical Jeans equation as given by Mamon & Łokas (2005). Finally, we calculate the integral over the spectroscopic slit of the seeing-convolved brightness-weighted LOS velocity dispersion $\sigma^{\rm P}$ (equation 20 of Suyu et al. 2010), which we then compare to the measurements to compute the likelihood of the kinematics data,

$$P(\sigma | \boldsymbol{\nu}, \boldsymbol{\pi}, \kappa_{\rm ext}, r_{\rm ani}) = \frac{1}{\sqrt{2\pi}\sigma_\sigma} \exp\left[-\frac{(\sigma^{\rm P}(\boldsymbol{\nu}, \boldsymbol{\pi}, \kappa_{\rm ext}, r_{\rm ani}) - \sigma)^2}{2\sigma_\sigma^2} \right], \tag{8}$$

where $\sigma = 222\,{\rm km\,s^{-1}}$ and $\sigma_\sigma = 15\,{\rm km\,s^{-1}}$ (Section 3.3). We adopt a uniform prior on $r_{\rm ani}$ between 0.5 and 5 times the effective radius,

---

[5] θ_γ is defined to be the direction of the shear itself, i.e. orthogonal to the direction of the mass producing the shear.

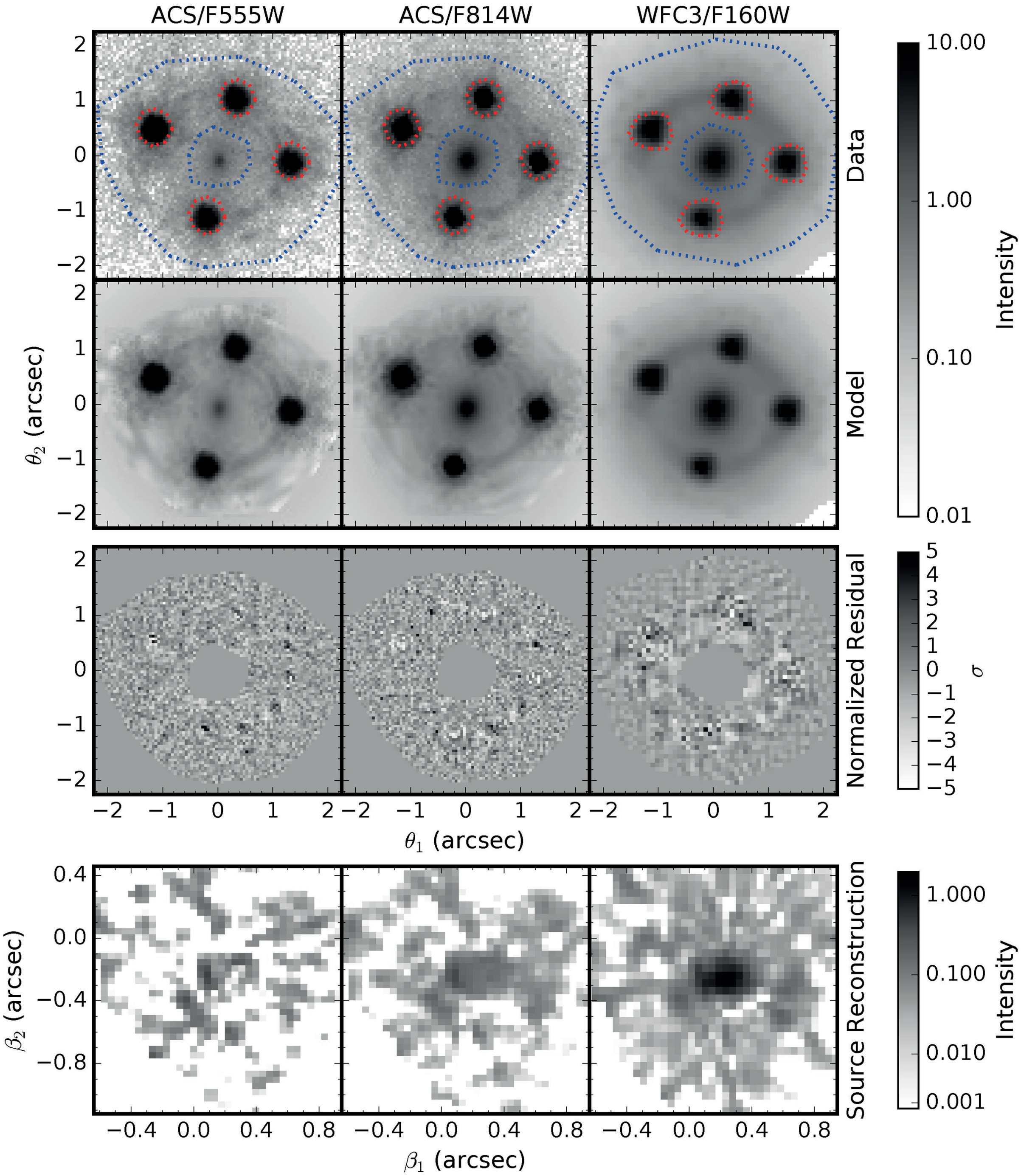


Figure 4. Lens model results for ACS/*F*555*W* (left), ACS/*F*814*W* (middle) and WFC3/*F*160*W* (right). Shown are the observed image (top row), the reconstructed image predicted by the model (second row), the normalized residual within the arcmask region (defined as the difference between the data and model, normalized by the estimated uncertainty of each pixel; third row) and the reconstructed source (bottom row). The blue dotted lines indicate the arcmask region used for fitting the extended source, and the red dotted lines indicate the AGN mask region where the power-law weighting is applied. The colour bars show the scale in the respective panels. The results shown here are for the fiducial model, but the results for the other systematics tests (Section 4.7) are qualitatively similar.

$r_{\rm eff}$, which we determine to be $r_{\rm eff} = 1.33$ arcsec from our lens light fitting[6] in the $F814W$ filter.

4.4 External convergence

In H0LiCOW Paper III, we estimate the external convergence using weighted number counts in a manner similar to Greene et al. (2013, see also Fassnacht et al. 2011). We use the weighted counts to select corresponding lines of sight from the $\kappa_{\rm ext}$ catalogues produced from the Millennium Simulation by Hilbert et al. (2009) and, thus, to get a $\kappa_{\rm ext}$ distribution. We use the $\kappa_{\rm ext}$ distribution from H0LiCOW Paper III that was derived by combining three constraints: the unweighted galaxy number counts, the counts weighted by $1/r$, and external shear matching that from our lens modelling, which gives a median external convergence at the position of HE 0435−1223 of $\kappa_{\rm ext} = 0.003$, with 16 per cent and 84 per cent percentiles of $\kappa_{\rm ext} = -0.016$ and $\kappa_{\rm ext} = 0.034$, respectively. Although the external shear can change slightly among different models, these changes generally affect the $\kappa_{\rm ext}$ distribution at the ∼0.005 level or smaller, which we can safely neglect. Since we are explicitly including the nearest LOS perturber in our mass model, this galaxy does not contribute to the inferred external shear, nor do we want it to be double counted in the external convergence. We therefore exclude galaxies projected within 5 arcsec of the main lens galaxy when calculating the relative galaxy number counts[7] for both the simulated and real lines of sight.

The host galaxy group likely has a small effect as the external shear is small, and an estimate of its flexion shift (H0LiCOW Paper II) indicates that it is a less significant perturber than G1. In addition, a weak lensing analysis of the field (Tihhonova et al., in preparation) finds a conservative 3σ upper limit of $\kappa_{\rm ext} = 0.04$ at the lens position, further suggesting that the group does not significantly affect our analysis. The external convergence contribution of the host galaxy group is implicitly included in our model through the procedure of H0LiCOW Paper III.

4.5 Blind analysis

Throughout our analysis, we blind the H_0 values in our lens model and the inferred time-delay distance values to avoid confirmation bias using a similar procedure as Suyu et al. (2013). This is done by subtracting the median of the parameter PDFs from the distribution when displaying plots. This allows us to measure the precision and relative offsets of these parameter distributions and their correlation with other parameters without being able to see the absolute value. This eliminates the tendency for experimenters to stop investigating systematic errors when they obtain an answer consistent with the 'expected' result. After finalizing our analysis, writing our paper draft with blinded $D_{\Delta t}$ distributions, and coming to a consensus among the coauthors during a collaboration telecon on 2016 June 16, we unblind the results and do not make any further changes to the models. There is also no iteration between the lens modelling and time delay measurements (i.e. the delays are measured once and used as they are; see H0LiCOW Paper V). Throughout this paper, we show blinded $D_{\Delta t}$ distributions until Section 5, where we reveal the absolute $D_{\Delta t}$ values from our inference.

4.6 Inferring the time-delay distance

Our inference on $D_{\Delta t}$ using all of the available data is calculated as in equations (4) and (5). We use importance sampling (e.g. Lewis & Bridle 2002) to combine the velocity dispersion and external convergence distributions with the $D_{\Delta t}^{\rm model}$ inferred from our lens model. Specifically, for each set of lens parameters $\boldsymbol{\nu}$ from our lens mass model MCMC chain, we draw a sample of $\kappa_{\rm ext}$ from the distribution in Section 4.4 and a sample of $r_{\rm ani}$ from the uniform distribution $[0.5,5]r_{\rm eff}$. With these, we can then compute the kinematics likelihood in equation (8) for the joint sample $\{\boldsymbol{\nu}, \kappa_{\rm ext}, r_{\rm ani}\}$ and use this to weight the joint sample. From the effective *model* time-delay distance computed from our multiplane lensing ($D_{\Delta t}^{\rm model}$) and the external convergence ($\kappa_{\rm ext}$), we can then compute the effective time-delay distance ($D_{\Delta t}$) via equation (3), keeping its absolute value blinded until we finalize our analysis. The resulting distribution of $D_{\Delta t}$ encapsulates the cosmological information from HE 0435−1223.

4.7 Systematics tests

In this section, we describe a range of tests of the effects of various systematics in our modelling. In addition to a basic 'fiducial' model, we perform inferences given the following sets of assumptions:

(i) a model with the image plane cutout region in all bands increased by 10 pixels in both the θ_1 and θ_2 directions.

(ii) A model with the arcmask region increased by 1 pixel on both the inner and outer edges. To compensate for the larger arcmask region, we increase the source plane resolution to 50 × 50 pixels in all bands.

(iii) A model with the arcmask region increased by 2 pixels on both the inner and outer edges. To accommodate the larger arcmask, we also increase the image plane cutout region by 10 pixels in all bands. To compensate for the larger arcmask region, we increase the source plane resolution to 50 × 50 pixels in all bands.

(iv) A model where the regions near the AGN images are given zero weight rather than being scaled by a power-law weighting.

(v) A model where the regions near the AGN images scaled by the power-law weighting is increased by 1 pixel around the outer edge.

(vi) A model where the regions near the AGN images scaled by the power-law weighting is increased by 2 pixels around the outer edge.

(vii) A model where the light profile of the lens galaxy is represented by the sum of two Sérsic profiles rather than the sum of two Chameleon profiles.

(viii) A model including the five most significant nearby perturbers (shown in Fig. 3) rather than just G1. The relative *Einstein* radii of the perturbers, assumed to be singular isothermal spheres, are calculated from their stellar masses (H0LiCOW Paper III), assuming a relationship between velocity dispersion and stellar mass from Bernardi et al. (2011). The ratio of *Einstein* radii is fixed, but with a global scaling allowed to vary freely. This is done to prevent the model from optimizing the perturbers' *Einstein* radii in a way that would be inconsistent with their measured redshifts and stellar masses. The galaxies' stellar masses are computed assuming the cosmology of the Millennium Simulation ($H_0 = 73\,{\rm km\,s^{-1}\,Mpc^{-1}}$,

---

[6] We use the double Sérsic model of the lens galaxy light to determine $r_{\rm eff}$ because the Chameleon profile does not provide an accurate description at large radii.

[7] For our model that includes the five nearest perturbers, we run a test where we calculate $\kappa_{\rm ext}$ excluding a larger region. The corresponding shift in $\kappa_{\rm ext}$ affects our final $D_{\Delta t}$ distribution by ∼0.2 per cent at most, so we neglect this effect.

$\Omega_m = 0.25$, $\Omega_\Lambda = 0.75$; Springel et al. 2005; Hilbert et al. 2009), but we verify that for alternative cosmologies, their stellar masses change by <0.02 dex, and the ratios of their *Einstein* radii therefore are affected by a negligible amount.

(ix) A ‘composite’ model with separate stellar and dark matter components. The details of this model are discussed in Section 4.8.

(x) The composite model with the regions near the AGN images scaled by the power-law weighting increased by 1 pixel around the outer edge.

(xi) The composite model with the arcmask region increased by 1 pixel on both the inner and outer edges and a 50 × 50 pixel source plane resolution.

As described in Section 2.2, we combine the MCMC chains from all of these tests. In doing so, we effectively assume that (1) these various tests sample a reasonable distribution of assumptions that we could have made when modelling the system, and that these assumptions have equal prior probability, and (2) neither the goodness of fit nor the parameter space prior volume are appreciably different between the tests. We verify that the goodness of fit does not change appreciably during this procedure (see Section 5). We weight the different MCMC chains equally and concatenate them, resulting in a set of samples that characterizes our final posterior PDF for $D_{\Delta t}$. This procedure folds the systematic uncertainty due to our modelling assumptions into our final uncertainty on the inferred parameters.

4.8 Comparison of power-law and composite models

We follow Suyu et al. (2014) to construct the composite model of baryons and dark matter as one of our systematics tests. The composite model consists of mass components associated with each of the four non-singular isothermal elliptical profiles (making up the two Chameleon profiles) in the lens galaxy light model in the WFC3/*F*160*W* band scaled by an overall mass-to-light (*M*/*L*) ratio. We use the *F*160*W* band because it probes the rest-frame near-infrared and thus should be the best tracer of stellar mass. The dark matter component is modelled as an elliptical NFW (Navarro, Frenk & White 1996) potential with its centroid linked to that of the light centroid in *F*160*W*. This is motivated by Dutton & Treu (2014), who find that non-contracted NFW profiles are a good representation for the dark matter haloes of massive elliptical galaxies.

The composite model has the following free parameters:

(i) *M*/*L* ratio for the baryonic component;

(ii) NFW halo normalization $\kappa_{0,h}$ (defined as $\kappa_{0,h} \equiv 4\kappa_s$; Golse & Kneib 2002);

(iii) NFW halo scale radius r_s;

(iv) NFW halo minor-to-major axis ratio q and associated position angle θ_q;

(v) external shear $\gamma_{\rm ext}$ and associated position angle θ_γ;

(vi) *Einstein* radius of G1;

(vii) the cosmological parameter H_0.

We set a Gaussian prior of $r_s = 14.3$ arcsec ± 2.0 arcsec based on the results of Gavazzi et al. (2007) for lenses in the Sloan Lens ACS Survey (SLACS; Bolton et al. 2006) sample, which encompasses the redshift and stellar mass of HE 0435−1223. All other parameters are given uniform priors. We note that the relative amplitudes of the two Chameleon profiles representing the stellar light distribution of the lens galaxy can vary during the modelling, whilst the relative amplitudes are fixed in the mass profiles. To account for this, we adopt an iterative approach where we run a series of MCMC chains and update the (fixed) relative amplitudes of the associated mass components to match that of the light components after each chain. We iterate until the relative change in the light profile amplitudes reach a point where the inferred $D_{\Delta t}$ stabilizes, then combine the MCMC chains after this point into a single distribution to represent the composite model. The remaining two composite models (with a larger arcmask or AGN mask) use fixed relative amplitudes of the mass components from the latest iteration of the original composite model.

The marginalized parameter distributions of the SPEMD model are shown in Fig. 5. We show the combined distributions of all SPEMD models as well as the fiducial model separately. The parameter statistics for each model are given in Appendix A. We note that two particular models stand out. The model with the arcmask expanded by 1 pixel and a 50 × 50 source grid prefers a smaller *Einstein* radius for the main lens galaxy and a larger *Einstein* radius for G1. This degeneracy is likely due to systematics associated with the source pixel size (Suyu et al. 2013), as this model has a smaller source pixel size than the others. The 5-perturber model prefers a smaller *Einstein* radius for both the main lens galaxy and G1, as well as a very different θ_γ. This is not surprising, as the addition of the extra perturbers in the lens model contributes to the integrated LOS lensing effect, reducing the contribution needed from the main lens and G1, as well as changing the external shear needed to fit the data. The offset between the mass centroid and the light centroid in the *F*160*W* band is typically ∼0.002 arcsec.

We show the marginalized parameter distributions of the composite model in Fig. 6. Again, we show the combined distributions as well as the main composite model separately, and the parameter statistics for each model are given in Appendix A. The main composite model appears to have some degenerate or bimodal features, but this is because this model itself is the combination of several separate models with slightly different relative amplitudes between the two Chameleon components, as mentioned above. The model with a larger arcmask and source grid prefers a larger G1 *Einstein* radius, similar to the analogous SPEMD model. The dark matter fraction within the *Einstein* radius for the composite models is $f_{\rm DM} \sim 45$ per cent.

We compare the physical parameters of our ‘fiducial’ power-law model to the composite model. The results are shown in Table 1, with the parameter statistics for all composite models given in Appendix A. We note that the external shear strength of the composite model is smaller than that of the power-law model, which we attribute to a degeneracy between $\gamma_{\rm ext}$ and the internal ellipticity of the mass model. When external shear is removed, the composite model's critical curves appear slightly more elliptical than those of the power-law model, supporting this interpretation. As mentioned in Section 4.4, the difference in $\gamma_{\rm ext}$ between these models has a negligible effect on the $\kappa_{\rm ext}$ distribution.

4.9 Impact of different cosmologies

In the multilens-plane modelling, we need to sample the cosmological parameters in order to carry out the ray tracing. Throughout our analysis, we only vary H_0, keeping other cosmological parameters fixed ($\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$ and $w = -1$). This is done for computational reasons, as the MCMC sampling becomes inefficient when they are all allowed to vary simultaneously. In principle, $D_{\Delta t}$ has a weak dependence on these other cosmological parameters. We test their impact by rerunning the fiducial model while allowing combinations of them to vary with uniform priors. The resulting effective $D_{\Delta t}$ distributions, shown in Fig. 7, have peaks that are consistent

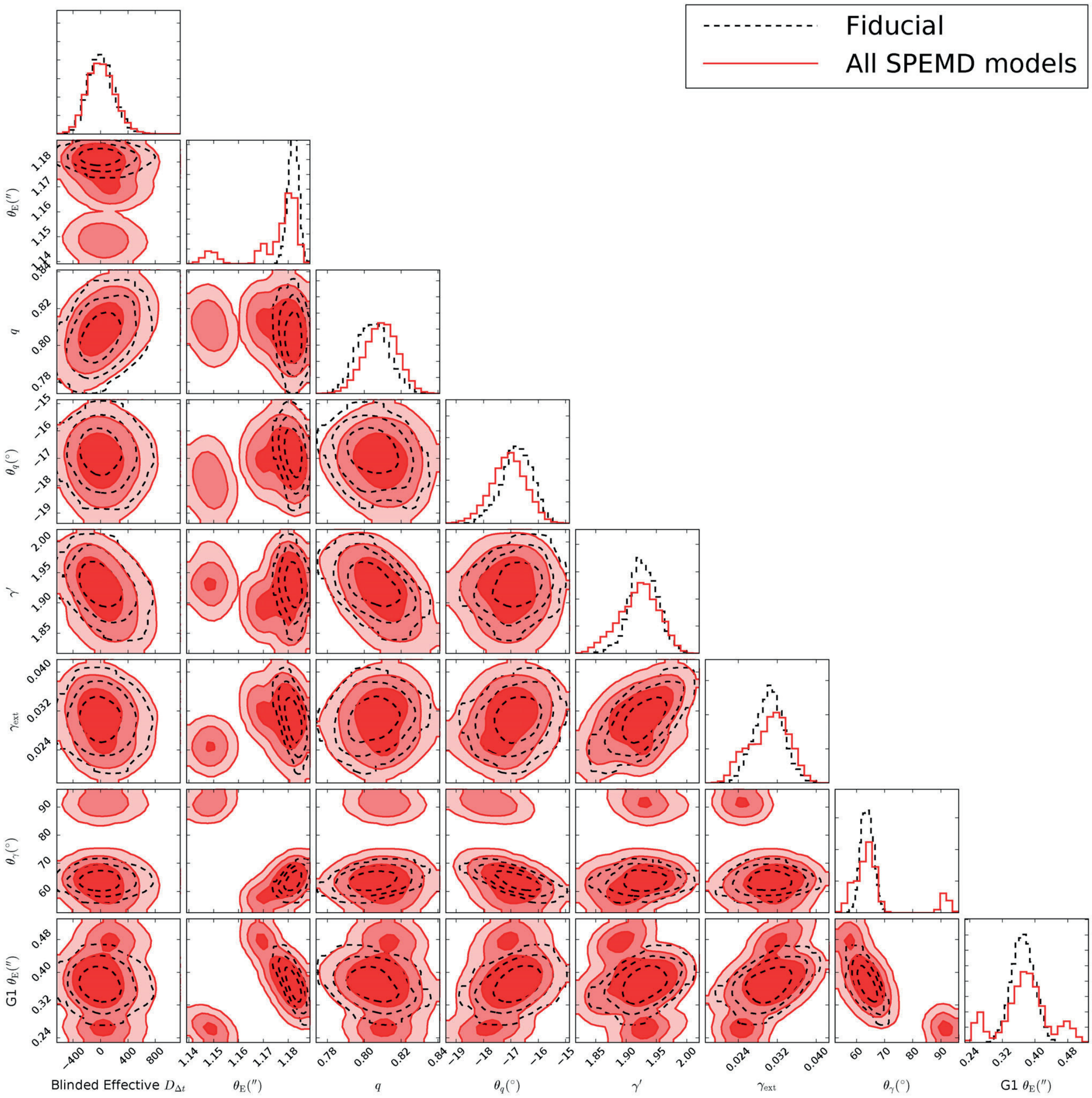


Figure 5. Marginalized parameter distributions from our SPEMD lens model results. We show the fiducial model (dashed black contours) and the combined results from our systematics tests (shaded red contours). The contours represent the 68.3 per cent, 95.4 per cent and 99.7 per cent quantiles.

to within 1 per cent of the absolute value, demonstrating that the results are insensitive to these extra cosmological parameters at the level of accuracy that we are currently working at. In future, when errors shrink further, this sampling will be included. We conclude that with the current level of precision, we are justified in deriving the posterior distribution function of the time-delay distance by varying H_0 only for computational efficiency. We emphasize that this does not affect in any way the generality of our results and that the resulting posterior distribution function is robust and can be interpreted in any cosmological model.

To expand on this point, it is instructive to consider multiplane lensing (e.g. Blandford & Narayan 1986; Kovner 1987; Kochanek & Apostolakis 1988; Schneider et al. 1992; Petters et al. 2001; Collett & Auger 2014; McCully et al. 2014; Schneider 2014) for the case of two lens planes, as we have in most of our models. Defining $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3$ as the angular coordinates on the main lens plane, the G1 lens plane and the source plane, respectively, the multiplane lens equations in this case are

$$\boldsymbol{\theta}_2 = \boldsymbol{\theta}_1 - \frac{D_{12}}{D_2}\hat{\boldsymbol{\alpha}}_1(D_1\boldsymbol{\theta}_1), \tag{9}$$

$$\boldsymbol{\theta}_3 = \boldsymbol{\theta}_1 - \frac{D_{13}}{D_3}\hat{\boldsymbol{\alpha}}_1(D_1\boldsymbol{\theta}_1) - \frac{D_{23}}{D_3}\hat{\boldsymbol{\alpha}}_2(D_2\boldsymbol{\theta}_2), \tag{10}$$

where D_i is the angular diameter distance from the observer to plane i, D_{ij} is the angular diameter distance between planes i and j and

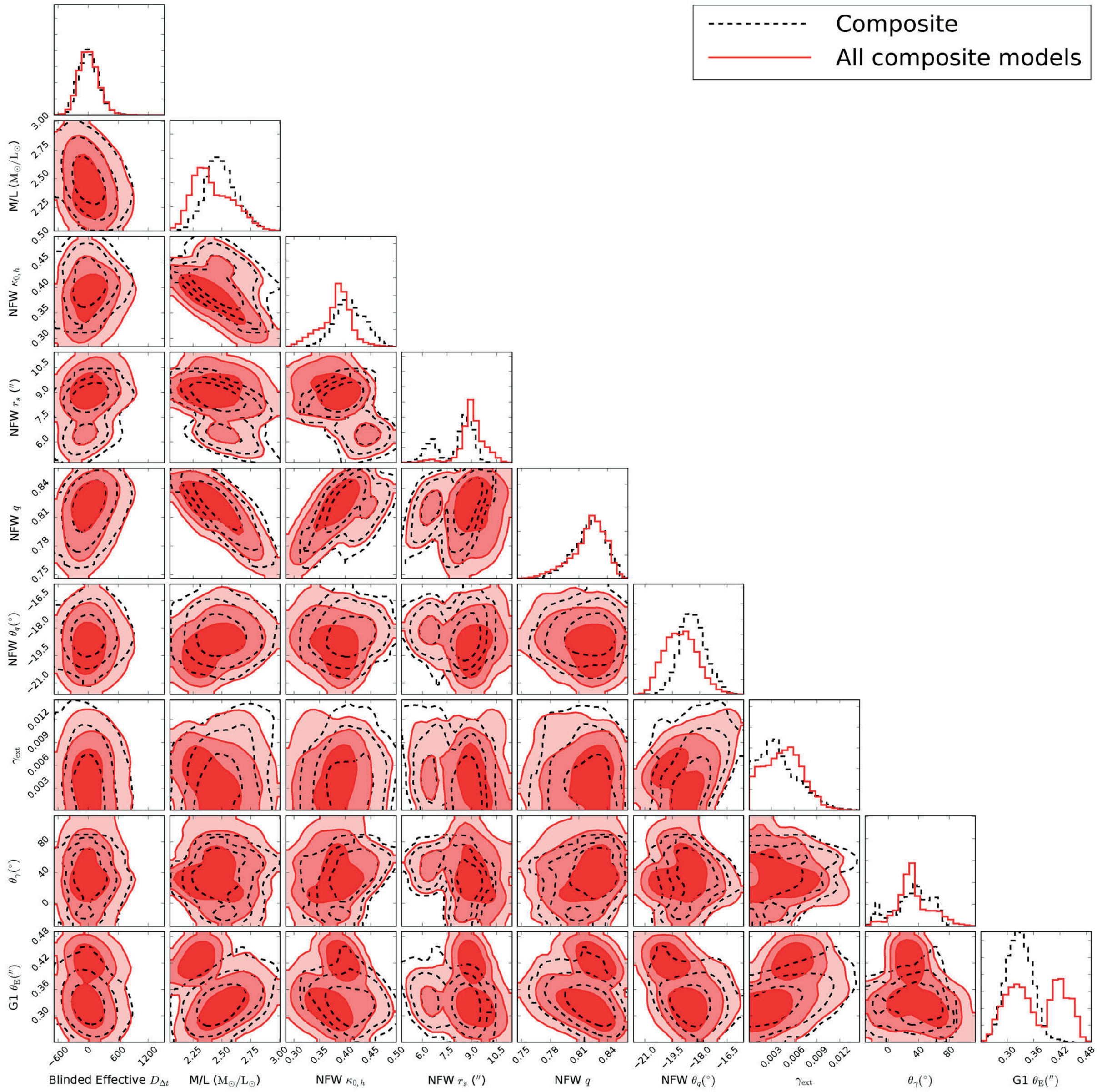


Figure 6. Marginalized parameter distributions from our composite lens model results. We show the main composite model (dashed black contours) and the combined results from our systematics tests (shaded red contours). The contours represent the 68.3 per cent, 95.4 per cent and 99.7 per cent quantiles.

$\hat{\boldsymbol{\alpha}}_i$ is the deflection angle at plane i. Scaling the deflection angle relative to the source (third) plane, we have the scaled deflection angle as

$$\boldsymbol{\alpha}_i(\boldsymbol{\theta}_i) = \frac{D_{i3}}{D_3}\hat{\boldsymbol{\alpha}}_i(D_i\boldsymbol{\theta}_i). \tag{11}$$

By further defining

$$\beta_{ij} = \frac{D_{ij}}{D_j}\frac{D_3}{D_{i3}}, \tag{12}$$

we can rewrite equations (9) and (10) as

$$\boldsymbol{\theta}_2 = \boldsymbol{\theta}_1 - \beta_{12}\boldsymbol{\alpha}_1(\boldsymbol{\theta}_1), \tag{13}$$

$$\boldsymbol{\theta}_3 = \boldsymbol{\theta}_1 - \boldsymbol{\alpha}_1(\boldsymbol{\theta}_1) - \boldsymbol{\alpha}_2(\boldsymbol{\theta}_2). \tag{14}$$

The multiplane time delay has contributions from the geometric delays between planes and the gravitational delay at each mass plane:

$$t = \frac{D_{\Delta t}^{12}}{c}\left[\frac{1}{2}|\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1|^2 - \beta_{12}\psi_1(\boldsymbol{\theta}_1)\right] + \frac{D_{\Delta t}^{23}}{c}\left[\frac{1}{2}|\boldsymbol{\theta}_3 - \boldsymbol{\theta}_2|^2 - \psi_2(\boldsymbol{\theta}_2)\right], \tag{15}$$

Table 1. Lens model parameters. Reported values are medians, with errors corresponding to the 16th and 84th percentiles. Angles are measured east of north.

Parameter	Marginalized constraints
Fiducial singular power-law ellipsoid model	
$\theta_{\rm E}$ (arcsec)[a]	$1.182^{+0.002}_{-0.002}$
q	$0.80^{+0.01}_{-0.01}$
θ_q (°)	$-16.8^{+0.5}_{-0.6}$
γ'	$1.93^{+0.02}_{-0.02}$
$\gamma_{\rm ext}$	$0.030^{+0.003}_{-0.003}$
θ_γ (°)	$63.7^{+2.4}_{-2.2}$
G1 $\theta_{\rm E}$ (arcsec)	$0.37^{+0.03}_{-0.03}$
Composite Model	
Stellar *M*/*L* ($\mathrm{M}_\odot/\mathrm{L}_\odot$)[b]	$2.5^{+0.1}_{-0.1}$
NFW $\kappa_{0,h}$	$0.41^{+0.03}_{-0.03}$
NFW r_s (arcsec)	$8.43^{+0.58}_{-1.94}$
NFW q	$0.82^{+0.01}_{-0.02}$
NFW θ_q (°)	$-18.4^{+0.7}_{-0.7}$
$\gamma_{\rm ext}$	$0.004^{+0.003}_{-0.002}$
θ_γ (°)	$34.4^{+22.9}_{-32.5}$
G1 $\theta_{\rm E}$ (arcsec)	$0.33^{+0.03}_{-0.03}$

Notes. [a]Spherical-equivalent *Einstein* radius.
[b]*M*/*L* within $\theta_{\rm E}$ for rest-frame *V* band. The given uncertainties are only statistical and do not include systematic effects. The stellar mass is calculated assuming $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_{\rm m} = 0.3$ and $\Omega_\Lambda = 0.7$, but changes in the cosmology affect the *M*/*L* by a negligible amount.

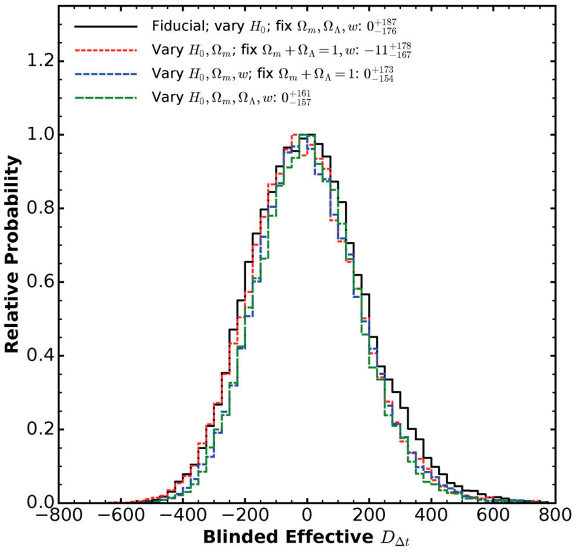


Figure 7. PDF of $D_{\Delta t}$ for the various cosmologies. We compare the fiducial model to one in which $\Omega_{\rm m}$ is allowed to vary (with $\Omega_{\rm m} + \Omega_\Lambda = 1$), one in which w is also allowed to vary, and one in which $\Omega_{\rm m}$, Ω_Λ and w are all allowed to vary independently. The distributions are blinded by subtracting the median of the fiducial model PDF. The different cosmology tests are indicated by the legend, and the median and 68 per cent quantiles of the $D_{\Delta t}$ distributions are given. The median of the blinded effective time-delay distance PSF is insensitive to the extra cosmological parameters to within 1 per cent.

Table 2. Effective time-delay distance. Reported values are medians, with errors corresponding to the 16th and 84th percentiles. χ^2 values are computed within the fiducial arcmask and outside the AGNmask+2pix region for a fair comparison among models. The models are grouped such that those that use the same data set are together. For models with a larger arcmask, we calculate χ^2 for a source grid resolution that approximately matches that of the other models so that we can fairly compare them.

Model	$D_{\Delta t}$ (Mpc)	χ^2
Fiducial	2532^{+187}_{-176}	11 024.9
Sérsic profiles	2722^{+209}_{-185}	11 001.5
5 perturbers	2642^{+187}_{-173}	11 002.0
Composite	2646^{+202}_{-188}	11 014.1
AGN mask+1pix	2507^{+189}_{-171}	11 029.6
Composite,AGN mask+1pix	2656^{+211}_{-194}	11 032.2
Arcmask+1pix,50x50src	2741^{+170}_{-150}	11 097.7
Composite,Arcmask+1pix,50x50src	2665^{+195}_{-171}	11 121.5
img+10pix	2532^{+184}_{-146}	11 090.2
Arcmask+2pix,img+10pix,50x50src	2636^{+178}_{-152}	11 074.3
AGN mask weight=0	2518^{+195}_{-181}	10 921.6
AGN mask+2pix	2528^{+187}_{-166}	11 065.4
Total	$\mathbf{2612^{+208}_{-191}}$	–

where ψ_i is the lens potential related to the scaled deflection angle via $\nabla\psi_i = \boldsymbol{\alpha}_i$, and the time-delay distances between planes are

$$D^{ij}_{\Delta t} \equiv (1 + z_i)\frac{D_i D_j}{D_{ij}}, \tag{16}$$

with z_i being the redshift of plane *i*. From equation (15), we see that the time delay depends on the two time-delay distances and β_{12}. In general, it is difficult to constrain all these distance quantities independently. In fact, in multiplane modelling, we adopt specific cosmological models to compute the distances (D_{ij} and D_i) for the ray tracing, and compare the time-delay distance measurements from these different background cosmologies. For the case of HE 0435−1223, where G1 is not strongly lensing the background source but merely perturbs it, the effect on the time delays from G1 is weak. The lack of sensitivity to $\Omega_{\rm m}$ and w suggests that HE 0435−1223 is not sensitive to β_{12} at an interesting level to probe it directly in the same way as a double source plane lens (e.g. Gavazzi et al. 2008; Collett & Auger 2014). In HE 0435−1223, we find instead that the time delays are mostly set by the strong lens, and we can measure the effective $D_{\Delta t}$, which is $D^{13}_{\Delta t}$, that is independent of assumptions on the background cosmology, as demonstrated in Fig. 7. This robust distance determination then permits us to constrain any reasonable cosmological model via the distance–redshift relation.

5 RESULTS

The marginalized posterior $D_{\Delta t}$ distributions for our lens model are given in Table 2. We report the median and 68 per cent quantiles for each of the models described in Section 4.7, as well as a final distribution that combines all of the chains. These distributions are shown in Fig. 8. The blinded distributions, shown on the bottom *x*-axis of Fig. 8, were the only values seen until the unblinding. The velocity dispersion and external convergence have

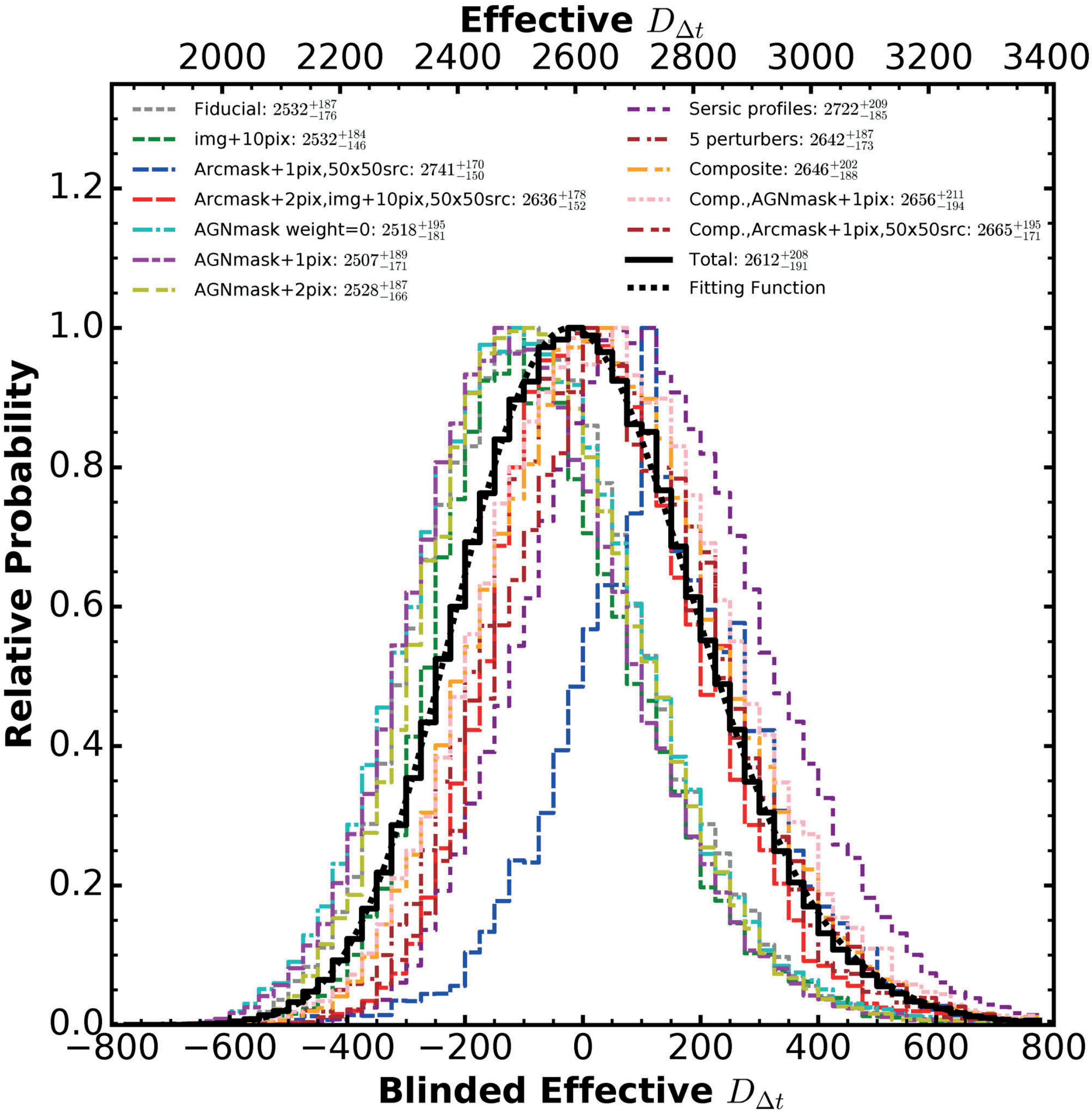


Figure 8. PDF of $D_{\Delta t}$ for the various models, as indicated by the legend. The median and 68 per cent quantile of each distribution is given. The thick black line represents the sum of all the distributions, which accounts for the various systematic uncertainties. The dotted black line is the skewed lognormal distribution (equation 17) fit to the final distribution. The bottom x-axis shows the blinded result, which is obtained by subtracting the median of the combined PDF from the absolute $D_{\Delta t}$ values. The top x-axis shows the true $D_{\Delta t}$ values. Throughout our blind analysis, the top x-axis was hidden until our analysis was finalized.

been included in these distributions. Each of the chains representing a different systematics test is given equal weight because the goodness-of-fit is comparable. Our final constraint on the effective time-delay distance in HE 0435−1223 is $D_{\Delta t} = 2612^{+208}_{-191}$ Mpc. We note that our fiducial model parameters are consistent with an identical model run only using the $F160W$ band as constraints.

Table 2 also shows the χ^2 for each model. The χ^2 values are calculated within the fiducial arcmask and outside of the AGN mask+2 pixel region to ensure a fair comparison among the different models. The χ^2 is calculated by summing the square of the normalized residual pixels (third row of Fig. 4) within this region. The number of degrees of freedom, $N_{\rm dof}$, is the number of pixels in this region across all three bands ($N_{\rm d} = 9577$) minus the number of lens mass/light model parameters minus a Γ term that represents the effective number of source pixels accounting for source regularization (see Suyu et al. 2006). Γ is calculated separately for each of the models' arcmask and AGN mask regions. The typical $N_{\rm dof}$ for our models is ∼8400–8600. Most of the residual χ^2 is associated with a few compact star-forming regions in the host galaxy that cannot be modelled at the resolution of our source pixel grid (Fig. 4). Our tests show that masking out these regions affects the $D_{\Delta t}$ distribution by less than our systematic uncertainties (see Section 4.2). We

Table 3. Cosmological parameter constraints from HE 0435−1223. Reported values are medians, with errors corresponding to the 16th and 84th percentiles. *Planck* priors are the Planck Collaboration XIII (2016) chains from baseline high-L *Planck* power spectra and low-L temperature and LFI polarization (*plikHM_TT_lowTEB*).

Model name	Description	Parameter priors	Marginalized cosmological parameters
$\mathrm{UH_0}$	Flat ΛCDM cosmology, fixed Ω_Λ $\Omega_\mathrm{m} = 1 - \Omega_\Lambda = 0.32$	H_0 uniform in [0,150]	$H_0 = 74.3^{+6.0}_{-5.4}$
UΛCDM	Flat ΛCDM cosmology $\Omega_\mathrm{m} = 1 - \Omega_\Lambda$	H_0 uniform in [0,150] Ω_Λ uniform in [0,1]	$H_0 = 73.1^{+5.7}_{-6.0}$ $\Omega_\Lambda = 0.51^{+0.34}_{-0.34}$
oΛCDM + *Planck*	Non-flat ΛCDM cosmology $\Omega_\mathrm{k} = 1 - \Omega_\Lambda - \Omega_\mathrm{m}$	*Planck* prior for $\{H_0, \Omega_\Lambda, \Omega_\mathrm{m}\}$	$H_0 = 63.5^{+3.7}_{-3.7}$ $\Omega_\mathrm{m} = 0.35^{+0.04}_{-0.04}$ $\Omega_\Lambda = 0.66^{+0.03}_{-0.03}$ $\Omega_\mathrm{k} = -0.011^{+0.010}_{-0.011}$
wCDM + *Planck*	Flat wCDM cosmology $\Omega_\mathrm{m} = 1 - \Omega_\Lambda$	*Planck* prior for $\{H_0, \Omega_\Lambda, w\}$	$H_0 = 83.7^{+9.2}_{-9.0}$ $\Omega_\Lambda = 0.80^{+0.04}_{-0.05}$ $w = -1.52^{+0.27}_{-0.27}$

note that for a fair comparison, the χ^2 for models with larger arcmasks are calculated on a source plane pixel scale that gives them approximately the same source resolution as the other models (41 × 41 pixels for the arcmask+1 pixel models, 45 × 45 pixels for the arcmask+2 pixel model). The typical absolute change in χ^2 for 1 pixel changes[8] in the source grid resolution is ∼60–70. We take this as the uncertainty in χ^2, and the χ^2 values are all very close among models that use the same data set. Therefore, we are justified in weighting each of the models equally.

We fit a skewed lognormal function to the $D_{\Delta t}$ distribution, as this function provides an accurate parametrized representation of our result (Suyu et al. 2010). The distribution has the form

$$P(D_{\Delta t}) = \frac{1}{\sqrt{2\pi}(x - \lambda_\mathrm{D})\sigma_\mathrm{D}} \exp\left[-\frac{(\ln(x - \lambda_\mathrm{D}) - \mu_\mathrm{D})^2}{2\sigma_\mathrm{D}^2}\right], \quad (17)$$

where $x = D_{\Delta t}/(1\,\mathrm{Mpc})$, $\lambda_\mathrm{D} = 653.9$, $\mu_\mathrm{D} = 7.5793$ and $\sigma_\mathrm{D} = 0.103\,12$. We plot this best-fitting function along with the final $D_{\Delta t}$ distribution in Fig. 8. The median, 68 per cent, and 95 per cent quantiles of the $D_{\Delta t}$ distribution and the best-fitting function agree to within ∼0.1 per cent, indicating that this function is an accurate representation.

Based on our inferred effective time-delay distance, we can calculate cosmological parameters for a variety of cosmological models, which are described in Table 3. For the UΛCDM cosmology, we constrain the Hubble constant to be $H_0 = 73.1^{+5.7}_{-6.0}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, giving a precision of ∼8 per cent from just this single lens system. This value is in good agreement with the latest distance ladder results ($H_0 = 73.24 \pm 1.74\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$; Riess et al. 2016) and higher than the latest *Planck* measurement for a similar cosmology ($H_0 = 67.8 \pm 0.9\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$; Planck Collaboration XIII 2016). Fig. 9 shows the posterior distribution of H_0 and Ω_m in UΛCDM. Fixing Ω_Λ in the $\mathrm{UH_0}$ model does not change the inferred H_0 significantly ($H_0 = 74.3^{+6.0}_{-5.4}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$). Our results for the oΛCDM + *Planck* model suggest a Universe consistent with spatial flatness. Interestingly, the wCDM + *Planck* model prefers a dark energy equation of state parameter that is in mild tension with $w = -1$ at the ∼2σ level. The results for each of our models are summarized in Table 3.

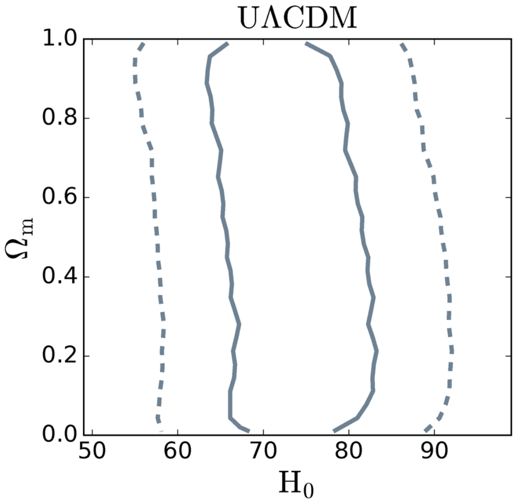


Figure 9. Posterior distribution of H_0 and Ω_m for the UΛCDM cosmology determined from the time-delay distance inference of HE 0435−1223. The contours represent the 68 per cent and 95 per cent quantiles of the distribution. Ω_m has a weak influence on $D_{\Delta t}$, so it is not well constrained. The marginalized value of H_0 for this cosmology is $73.1^{+5.7}_{-6.0}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

The results for HE 0435−1223 presented here can be combined with previous analyses of B1608+656 (Suyu et al. 2010) and RXJ1131−1231 (Suyu et al. 2013, 2014) to produce stronger constraints on cosmology. A full analysis of the implications of our $D_{\Delta t}$ inference for a variety of cosmologies using constraints from all three H0LiCOW lenses analysed to date is presented in H0LiCOW Paper V.

6 CONCLUSIONS

We have performed a blind analysis of the gravitational lens HE 0435−1223 using new deep *HST* imaging, high-precision time delays from COSMOGRAIL, a measurement of the lens galaxy velocity dispersion, and spectroscopic and photometric data to constrain the mass distribution along the LOS. Our model is able to reproduce the surface brightness structure of the lensed AGN and

---

[8] A 1 pixel change in source grid resolution roughly corresponds to the changes in source pixel size across our different models.

host galaxy in all three *HST* bands, as well as the measured time delays. Combining these data sets and accounting for various sources of systematic uncertainty in the lens modelling, we constrain the effective time-delay distance to be $D_{\Delta t} = 2612^{+208}_{-191}$ Mpc, giving a precision of 7.6 per cent. For a flat ΛCDM cosmology with uniform priors on H_0 and Ω_Λ, we constrain the Hubble constant to be $H_0 = 73.1^{+5.7}_{-6.0}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (a precision of ~8 per cent), in good agreement with the latest distance ladder results. A detailed analysis of the implications of our $D_{\Delta t}$ constraint on a variety of cosmologies is presented in H0LiCOW Paper V.

Upcoming analyses of the remaining two H0LiCOW systems will complete the sample of five time-delay lenses and constrain H_0 to <3.5 per cent precision. Our extensive blind analysis of HE 0435−1223 demonstrates the utility of gravitational lens time delays as a precise and independent cosmological probe. With hundreds of new lensed AGN expected to be discovered in current and future wide-field imaging surveys (Oguri & Marshall 2010), we expect time-delay cosmography to provide competitive cosmological constraints throughout the next decade.

ACKNOWLEDGEMENTS

We thank the referee for reviewing this paper and providing helpful commentary. We thank Adriano Agnello, Roger Blandford, Geoff Chih-Fan Chen, Xuheng Ding, Yashar Hezaveh, Kai Liao, John McKean, Georges Meylan, Danka Paraficz, Chiara Spiniello, Malte Tewes, Olga Tihhonova and Simona Vegetti for their contributions to the H0LiCOW project. We thank Simon Birrer for useful discussions and feedback. We thank Bau-Ching Hsieh for computing support on the SuMIRe computing cluster. We thank Michelle Wilson for sharing the lens galaxy group properties from an ongoing independent study. H0LiCOW and COSMOGRAIL are made possible thanks to the continuous work of all observers and technical staff obtaining the monitoring observations, in particular at the Swiss Euler telescope at La Silla Observatory. Euler is supported by the Swiss National Science Foundation. KCW is supported by an EACOA Fellowship awarded by the East Asia Core Observatories Association, which consists of the Academia Sinica Institute of Astronomy and Astrophysics, the National Astronomical Observatory of Japan, the National Astronomical Observatories of the Chinese Academy of Sciences, and the Korea Astronomy and Space Science Institute. SHS is supported by the Max Planck Society through the Max Planck Research Group. This work is supported in part by the Ministry of Science and Technology in Taiwan via grant MOST-103-2112-M-001-003-MY3. VB and FC are supported by the Swiss National Science Foundation (SNSF). CDF and CER are funded through the NSF grant AST-1312329, 'Collaborative Research: Accurate cosmology with strong gravitational lens time delays' and the *HST* grant GO-12889. DS acknowledges funding support from a *Back to Belgium* grant from the Belgian Federal Science Policy (BELSPO). TT thanks the Packard Foundation for generous support through a Packard Research Fellowship, the NSF for funding through NSF grant AST-1450141, 'Collaborative Research: Accurate cosmology with strong gravitational lens time delays'. SH acknowledges support by the DFG cluster of excellence 'Origin and Structure of the Universe' (www.universe-cluster.de). LVEK is supported in part through an NWO-VICI career grant (project number 639.043.308). PJM acknowledges support from the U.S. Department of Energy under contract number DE-AC02-76SF00515. This paper is based on observations made with the NASA/ESA *HST*, obtained at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. These observations are associated with Programs #12889 and #9744. Support for program #12889 was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. Some of the data presented herein were obtained at the W.M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California and the National Aeronautics and Space Administration. The Observatory was made possible by the generous financial support of the W.M. Keck Foundation. The authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Mauna Kea has always had within the indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this mountain.

REFERENCES

Anderson J., Bedin L. R., 2010, PASP, 122, 1035
Appenzeller I. et al., 1998, The Messenger, 94, 1
Auger M. W., Treu T., Bolton A. S., Gavazzi R., Koopmans L. V. E., Marshall P. J., Moustakas L. A., Burles S., 2010, ApJ, 724, 511
Barkana R., 1998, ApJ, 502, 531
Barnabè M., Czoske O., Koopmans L. V. E., Treu T., Bolton A. S., 2011, MNRAS, 415, 2215
Bernardi M., Roche N., Shankar F., Sheth R. K., 2011, MNRAS, 412, 684
Birrer S., Amara A., Refregier A., 2016, J. Cosmology Astropart. Phys., 8, 020
Blandford R., Narayan R., 1986, ApJ, 310, 568
Bolton A. S., Burles S., Koopmans L. V. E., Treu T., Moustakas L. A., 2006, ApJ, 638, 703
Bonvin V. et al., 2016a, MNRAS, preprint (arXiv:1607.01790) (H0LiCOW Paper V)
Bonvin V., Tewes M., Courbin F., Kuntzer T., Sluse D., Meylan G., 2016b, A&A, 585, A88
Cantale N., Courbin F., Tewes M., Jablonka P., Meylan G., 2016a, A&A, 589, A81
Cantale N. et al., 2016b, A&A, 589, A82
Chabrier G., 2003, PASP, 115, 763
Chen G. C.-F. et al., 2016, MNRAS, 462, 3457
Coe D., Moustakas L. A., 2009, ApJ, 706, 45
Collett T. E., Auger M. W., 2014, MNRAS, 443, 969
Collett T. E., Cunnington S. D., 2016, MNRAS, 462, 3255
Collett T. E. et al., 2013, MNRAS, 432, 679
Courbin F., Eigenbrod A., Vuissoz C., Meylan G., Magain P., 2005, in Mellier Y., Meylan G., eds, Proc. IAU Symp. 225, Gravitational Lensing Impact on Cosmology. Cambridge Univ. Press, Cambridge, p. 297
Courbin F. et al., 2011, A&A, 536, A53
Dutton A. A., Treu T., 2014, MNRAS, 438, 3594
Dutton A. A. et al., 2011, MNRAS, 417, 1621
Eigenbrod A., Courbin F., Vuissoz C., Meylan G., Saha P., Dye S., 2005, A&A, 436, 25
Fadely R., Keeton C. R., Nakajima R., Bernstein G. M., 2010, ApJ, 711, 246
Falco E. E., Gorenstein M. V., Shapiro I. I., 1985, ApJ, 289, L1
Fassnacht C. D., Pearson T. J., Readhead A. C. S., Browne I. W. A., Koopmans L. V. E., Myers S. T., Wilkinson P. N., 1999, ApJ, 527, 498
Fassnacht C. D., Xanthopoulos E., Koopmans L. V. E., Rusin D., 2002, ApJ, 581, 823
Fassnacht C. D., Gal R. R., Lubin L. M., McKean J. P., Squires G. K., Readhead A. C. S., 2006, ApJ, 642, 30
Fassnacht C. D., Koopmans L. V. E., Wong K. C., 2011, MNRAS, 410, 2167
Fazio G. G. et al., 2004, ApJS, 154, 10
Freedman W. L., Madore B. F., Scowcroft V., Burns C., Monson A., Persson S. E., Seibert M., Rigby J., 2012, ApJ, 758, 24

Gavazzi R., Treu T., Rhodes J. D., Koopmans L. V. E., Bolton A. S., Burles S., Massey R. J., Moustakas L. A., 2007, ApJ, 667, 176

Gavazzi R., Treu T., Koopmans L. V. E., Bolton A. S., Moustakas L. A., Burles S., Marshall P. J., 2008, ApJ, 677, 1046

Golse G., Kneib J.-P., 2002, A&A, 390, 821

Gorenstein M. V., Shapiro I. I., Falco E. E., 1988, ApJ, 327, 693

Greene Z. S. et al., 2013, ApJ, 768, 39

Hernquist L., 1990, ApJ, 356, 359

Hilbert S., Hartlap J., White S. D. M., Schneider P., 2009, A&A, 499, 31

Hook I. M., Jørgensen I., Allington-Smith J. R., Davies R. L., Metcalfe N., Murowinski R. G., Crampton D., 2004, PASP, 116, 425

Hu W., 2005, in Wolff S. C., Lauer T. R., eds, ASP Conf. Ser. Vol. 339, Observing Dark Energy. Astron. Soc. Pac., San Francisco, p. 215

Ilbert O. et al., 2006, A&A, 457, 841

Jaffe W., 1983, MNRAS, 202, 995

Kassiola A., Kovner I., 1993, ApJ, 417, 450

Keeton C. R., 2003, ApJ, 584, 664

Keeton C. R., Kochanek C. S., 1997, ApJ, 487, 42

Kochanek C. S., 2002, ApJ, 578, 25

Kochanek C. S., 2003, ApJ, 583, 49

Kochanek C. S., Apostolakis J., 1988, MNRAS, 235, 1073

Kochanek C. S., Morgan N. D., Falco E. E., McLeod B. A., Winn J. N., Dembicky J., Ketzeback B., 2006, ApJ, 640, 47

Koopmans L. V. E., 2004, preprint (astro-ph/0412596)

Koopmans L. V. E., Treu T., Fassnacht C. D., Blandford R. D., Surpi G., 2003, ApJ, 599, 70

Koopmans L. V. E., Treu T., Bolton A. S., Burles S., Moustakas L. A., 2006, ApJ, 649, 599

Koopmans L. V. E. et al., 2009, ApJ, 703, L51

Kovner I., 1987, ApJ, 316, 52

Lewis A., Bridle S., 2002, Phys. Rev. D, 66, 103511

McCully C., Keeton C. R., Wong K. C., Zabludoff A. I., 2014, MNRAS, 443, 3631

McCully C., Keeton C. R., Wong K. C., Zabludoff A. I., 2016, ApJ, preprint (arXiv:1601.05417)

Magain P., Courbin F., Sohy S., 1998, ApJ, 494, 472

Mamon G. A., Łokas E. L., 2005, MNRAS, 363, 705

Massey R., Stoughton C., Leauthaud A., Rhodes J., Koekemoer A., Ellis R., Shaghoulian E., 2010, MNRAS, 401, 371

Merritt D., 1985, AJ, 90, 1027

Momcheva I., Williams K., Keeton C., Zabludoff A., 2006, ApJ, 641, 169

Momcheva I. G., Williams K. A., Cool R. J., Keeton C. R., Zabludoff A. I., 2015, ApJS, 219, 29

Morgan N. D., Kochanek C. S., Pevunova O., Schechter P. L., 2005, AJ, 129, 2531

Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563

Oguri M., 2007, ApJ, 660, 1

Oguri M., Marshall P. J., 2010, MNRAS, 405, 2579

Oke J. B. et al., 1995, PASP, 107, 375

Osipkov L. P., 1979, Pis'ma Astron. Zh., 5, 77

Petters A. O., Levine H., Wambsganss J., 2001, Singularity Theory and Gravitational Lensing. Birkhäuser, Boston

Planck Collaboration XIII, 2016, A&A, 594, A13

Rathna Kumar S., Stalin C. S., Prabhu T. P., 2015, A&A, 580, A38

Refsdal S., 1964, MNRAS, 128, 307

Riess A. G. et al., 2011, ApJ, 730, 119

Riess A. G. et al., 2016, ApJ, 826, 56

Rusu C. E. et al., 2016, MNRAS, preprint (arXiv:1607.01047) (H0LiCOW Paper III)

Saha P., 2000, AJ, 120, 1654

Saha P., Coles J., Macciò A. V., Williams L. L. R., 2006, ApJ, 650, L17

Schechter P. L. et al., 1997, ApJ, 475, L85

Schneider P., 2014, A&A, 568, L2

Schneider P., Sluse D., 2013, A&A, 559, A37

Schneider P., Sluse D., 2014, A&A, 564, A103

Schneider P., Ehlers J., Falco E. E., 1992, Gravitational Lenses. Springer-Verlag, New York

Seljak U., 1994, ApJ, 436, 509

Sereno M., Paraficz D., 2014, MNRAS, 437, 600

Sluse D., Hutsemékers D., Courbin F., Meylan G., Wambsganss J., 2012, A&A, 544, A62

Sluse D. et al., 2016, MNRAS, preprint (arXiv:1607.00382) (H0LiCOW Paper II)

Sonnenfeld A., Treu T., Gavazzi R., Marshall P. J., Auger M. W., Suyu S. H., Koopmans L. V. E., Bolton A. S., 2012, ApJ, 752, 163

Sonnenfeld A., Treu T., Gavazzi R., Suyu S. H., Marshall P. J., Auger M. W., Nipoti C., 2013, ApJ, 777, 98

Springel V. et al., 2005, Nature, 435, 629

Suyu S. H., 2012, MNRAS, 426, 868

Suyu S. H., Halkola A., 2010, A&A, 524, A94

Suyu S. H., Marshall P. J., Hobson M. P., Blandford R. D., 2006, MNRAS, 371, 983

Suyu S. H., Marshall P. J., Blandford R. D., Fassnacht C. D., Koopmans L. V. E., McKean J. P., Treu T., 2009, ApJ, 691, 277

Suyu S. H., Marshall P. J., Auger M. W., Hilbert S., Blandford R. D., Koopmans L. V. E., Fassnacht C. D., Treu T., 2010, ApJ, 711, 201

Suyu S. H. et al., 2012a, preprint (arXiv:1202.4459)

Suyu S. H. et al., 2012b, ApJ, 750, 10

Suyu S. H. et al., 2013, ApJ, 766, 70

Suyu S. H. et al., 2014, ApJ, 788, L35

Suyu S. H. et al., 2016, MNRAS, preprint (arXiv:1607.00017) (H0LiCOW Paper I)

Tewes M., Courbin F., Meylan G., 2013a, A&A, 553, A120

Tewes M. et al., 2013b, A&A, 556, A22

Treu T., Koopmans L. V. E., 2002, ApJ, 575, 87

Treu T., Marshall P. J., 2016, A&AR, preprint (arXiv:1605.05333)

Unruh S., Schneider P., Sluse D., 2016, A&A, preprint (arXiv:1606.04321)

Vanderriest C., Schneider J., Herpe G., Chevreton M., Moles M., Wlerick G., 1989, A&A, 215, 1

Velander M. et al., 2014, MNRAS, 437, 2111

Warren S. J., Dye S., 2003, ApJ, 590, 673

Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, Phys. Rep., 530, 87

Williams K. A., Momcheva I., Keeton C. R., Zabludoff A. I., Lehár J., 2006, ApJ, 646, 85

Wisotzki L., Christlieb N., Bade N., Beckmann V., Köhler T., Vanelle C., Reimers D., 2000, A&A, 358, 77

Wisotzki L., Schechter P. L., Bradt H. V., Heinmüller J., Reimers D., 2002, A&A, 395, 17

Wong K. C., Keeton C. R., Williams K. A., Momcheva I. G., Zabludoff A. I., 2011, ApJ, 726, 84

Wucknitz O., 2002, MNRAS, 332, 951

Xu D., Sluse D., Schneider P., Springel V., Vogelsberger M., Nelson D., Hernquist L., 2016, MNRAS, 456, 739

APPENDIX A: MODEL PARAMETERS

We show the marginalized parameter constraints for each of the SPEMD models in Table A1 and for each of the composite models in Table A2.

Table A1. SPEMD model parameters. Reported values are medians, with errors corresponding to the 16th and 84th percentiles. Angles are measured east of north.

Parameter	Marginalized constraints								
	Fiducial	Img+10	Arc+1,50src	Arc+2,Im+10,50src	AGNwht=0	AGNmask+1	AGNmask+2	5 pert.	Sérsic
$\theta_{\rm E}$ (arcsec)[a]	$1.182^{+0.002}_{-0.002}$	$1.182^{+0.002}_{-0.002}$	$1.169^{+0.002}_{-0.001}$	$1.176^{+0.002}_{-0.002}$	$1.180^{+0.002}_{-0.002}$	$1.182^{+0.002}_{-0.002}$	$1.181^{+0.002}_{-0.002}$	$1.149^{+0.003}_{-0.003}$	$1.181^{+0.002}_{-0.002}$
q	$0.80^{+0.01}_{-0.01}$	$0.81^{+0.01}_{-0.01}$	$0.81^{+0.01}_{-0.01}$	$0.81^{+0.01}_{-0.01}$	$0.81^{+0.01}_{-0.01}$	$0.80^{+0.01}_{-0.01}$	$0.80^{+0.01}_{-0.01}$	$0.81^{+0.01}_{-0.01}$	$0.81^{+0.01}_{-0.01}$
θ_q (°)	$-16.8^{+0.5}_{-0.6}$	$-17.0^{+0.5}_{-0.5}$	$-17.1^{+0.5}_{-0.4}$	$-16.6^{+0.5}_{-0.6}$	$-16.8^{+0.5}_{-0.5}$	$-17.1^{+0.6}_{-0.6}$	$-17.3^{+0.5}_{-0.6}$	$-17.8^{+0.5}_{-0.5}$	$-17.0^{+0.6}_{-0.6}$
γ'	$1.93^{+0.02}_{-0.02}$	$1.95^{+0.02}_{-0.03}$	$1.89^{+0.02}_{-0.02}$	$1.91^{+0.02}_{-0.01}$	$1.94^{+0.02}_{-0.02}$	$1.94^{+0.02}_{-0.02}$	$1.94^{+0.02}_{-0.02}$	$1.93^{+0.02}_{-0.01}$	$1.87^{+0.03}_{-0.02}$
$\gamma_{\rm ext}$	$0.030^{+0.003}_{-0.003}$	$0.033^{+0.003}_{-0.003}$	$0.032^{+0.002}_{-0.002}$	$0.030^{+0.003}_{-0.004}$	$0.033^{+0.003}_{-0.003}$	$0.032^{+0.003}_{-0.003}$	$0.031^{+0.003}_{-0.003}$	$0.025^{+0.002}_{-0.002}$	$0.026^{+0.003}_{-0.003}$
θ_γ (°)	$63.7^{+2.4}_{-2.2}$	$65.0^{+1.9}_{-1.8}$	$57.7^{+1.2}_{-1.6}$	$60.6^{+2.1}_{-1.7}$	$63.6^{+1.9}_{-1.9}$	$65.3^{+1.9}_{-2.0}$	$65.4^{+2.0}_{-2.0}$	$-88.5^{+1.4}_{-1.3}$	$63.1^{+2.7}_{-2.7}$
G1 $\theta_{\rm E}$ (arcsec)	$0.37^{+0.03}_{-0.03}$	$0.38^{+0.02}_{-0.03}$	$0.48^{+0.02}_{-0.02}$	$0.40^{+0.02}_{-0.02}$	$0.39^{+0.02}_{-0.03}$	$0.37^{+0.03}_{-0.03}$	$0.37^{+0.03}_{-0.03}$	$0.26^{+0.01}_{-0.01}$	$0.35^{+0.03}_{-0.03}$

Note. [a]Spherical-equivalent *Einstein* radius.

Table A2. Composite model parameters. Reported values are medians, with errors corresponding to the 16th and 84th percentiles. Angles are measured east of north.

	Marginalized constraints		
Parameter	Composite	Composite, AGNmask+1	Composite, Arcmask+1,50src
Stellar *M*/*L* ($M_\odot/L_\odot$)[a]	$2.5^{+0.1}_{-0.1}$	$2.6^{+0.2}_{-0.2}$	$2.3^{+0.1}_{-0.1}$
NFW $\kappa_{0,h}$	$0.41^{+0.03}_{-0.03}$	$0.36^{+0.03}_{-0.03}$	$0.39^{+0.01}_{-0.01}$
NFW r_s (arcsec)	$8.43^{+0.58}_{-1.94}$	$9.43^{+0.69}_{-0.94}$	$8.96^{+0.28}_{-0.26}$
NFW q	$0.82^{+0.01}_{-0.02}$	$0.81^{+0.02}_{-0.02}$	$0.83^{+0.01}_{-0.01}$
NFW θ_q (°)	$-18.4^{+0.7}_{-0.7}$	$-18.6^{+0.7}_{-0.7}$	$-19.7^{+0.6}_{-0.6}$
$\gamma_{\rm ext}$	$0.004^{+0.003}_{-0.002}$	$0.003^{+0.003}_{-0.002}$	$0.006^{+0.002}_{-0.002}$
θ_γ (°)	$34.4^{+22.9}_{-32.5}$	$44.6^{+26.8}_{-36.2}$	$28.3^{+6.0}_{-7.6}$
G1 $\theta_{\rm E}$ (arcsec)	$0.33^{+0.03}_{-0.03}$	$0.32^{+0.03}_{-0.03}$	$0.42^{+0.03}_{-0.02}$

Note. [a]*M*/*L* within $\theta_{\rm E}$ for rest-frame *V* band. The given uncertainties are only statistical and do not include systematic effects. The stellar mass is calculated assuming $H_0 = 70\,{\rm km\,s^{-1}\,Mpc^{-1}}$, $\Omega_{\rm m} = 0.3$ and $\Omega_\Lambda = 0.7$, but changes in the cosmology affect the *M*/*L* by a negligible amount.

APPENDIX B: INVERSE MAGNIFICATION TENSORS

The components of the inverse magnification tensor are

$$\mathcal{A}_{ij}(\boldsymbol{\theta}) = \frac{\partial \beta_i}{\partial \theta_j}, \tag{B1}$$

where $i = 1, 2$, $j = 1, 2$, $\boldsymbol{\beta} = (\beta_1, \beta_2)$ is the source plane coordinates, and $\boldsymbol{\theta} = (\theta_1, \theta_2)$ is the coordinates of the image plane [which is also the first lens plane, $\boldsymbol{\theta} = (\theta_{1_1}, \theta_{1_2})$].

The general multiplane lens equation is

$$\boldsymbol{\theta}_j = \boldsymbol{\theta}_1 - \sum_{i=1}^{j-1} \beta_{ij} \boldsymbol{\alpha}_i(\boldsymbol{\theta}_i), \tag{B2}$$

where β_{ij} is given by equation (12) [note the difference between β_{ij} with two indices and the source coordinates β_i with one index]. This is the general form of equations (13) and (14). For N lens planes, the source coordinates are $\boldsymbol{\beta} = \boldsymbol{\theta}_{N+1}$. For the case of two lens planes, as we have in our model, $\boldsymbol{\beta} = \boldsymbol{\theta}_3$. We present the inverse magnification tensors at the positions of the lensed quasar images in Table B1. While the inverse magnification tensor is symmetric for single-plane lensing, this is not true for multiplane lensing.

Table B1. Inverse magnification tensor.

Model	A	B	C	D
Fiducial	$\begin{bmatrix} 0.732^{+0.018}_{-0.016} & -0.287^{+0.006}_{-0.007} \\ -0.302^{+0.007}_{-0.007} & 0.328^{+0.008}_{-0.008} \end{bmatrix}$	$\begin{bmatrix} -0.105^{+0.004}_{-0.004} & 0.252^{+0.007}_{-0.006} \\ 0.281^{+0.008}_{-0.007} & 0.857^{+0.023}_{-0.021} \end{bmatrix}$	$\begin{bmatrix} 0.832^{+0.020}_{-0.017} & -0.009^{+0.002}_{-0.002} \\ -0.024^{+0.003}_{-0.003} & 0.184^{+0.005}_{-0.005} \end{bmatrix}$	$\begin{bmatrix} -0.220^{+0.008}_{-0.007} & 0.212^{+0.007}_{-0.006} \\ 0.249^{+0.008}_{-0.007} & 0.850^{+0.025}_{-0.022} \end{bmatrix}$
img+10pix	$\begin{bmatrix} 0.743^{+0.014}_{-0.020} & -0.292^{+0.007}_{-0.006} \\ -0.308^{+0.009}_{-0.007} & 0.333^{+0.008}_{-0.010} \end{bmatrix}$	$\begin{bmatrix} -0.105^{+0.005}_{-0.004} & 0.256^{+0.006}_{-0.007} \\ 0.286^{+0.007}_{-0.009} & 0.876^{+0.020}_{-0.026} \end{bmatrix}$	$\begin{bmatrix} 0.846^{+0.015}_{-0.022} & -0.009^{+0.002}_{-0.002} \\ -0.024^{+0.003}_{-0.003} & 0.186^{+0.004}_{-0.006} \end{bmatrix}$	$\begin{bmatrix} -0.220^{+0.009}_{-0.006} & 0.216^{+0.006}_{-0.007} \\ 0.254^{+0.007}_{-0.009} & 0.870^{+0.021}_{-0.027} \end{bmatrix}$
Arcmask+1pix,50x50src	$\begin{bmatrix} 0.692^{+0.013}_{-0.013} & -0.274^{+0.004}_{-0.005} \\ -0.292^{+0.005}_{-0.006} & 0.313^{+0.006}_{-0.006} \end{bmatrix}$	$\begin{bmatrix} -0.099^{+0.003}_{-0.003} & 0.236^{+0.005}_{-0.005} \\ 0.271^{+0.007}_{-0.006} & 0.809^{+0.018}_{-0.016} \end{bmatrix}$	$\begin{bmatrix} 0.776^{+0.014}_{-0.014} & -0.010^{+0.002}_{-0.002} \\ -0.029^{+0.002}_{-0.002} & 0.178^{+0.004}_{-0.004} \end{bmatrix}$	$\begin{bmatrix} -0.207^{+0.005}_{-0.006} & 0.185^{+0.004}_{-0.004} \\ 0.230^{+0.006}_{-0.005} & 0.797^{+0.019}_{-0.017} \end{bmatrix}$
Arcmask+2pix,img+10pix,50x50src	$\begin{bmatrix} 0.711^{+0.011}_{-0.011} & -0.280^{+0.005}_{-0.005} \\ -0.296^{+0.005}_{-0.005} & 0.319^{+0.005}_{-0.005} \end{bmatrix}$	$\begin{bmatrix} -0.102^{+0.003}_{-0.003} & 0.243^{+0.005}_{-0.004} \\ 0.274^{+0.005}_{-0.005} & 0.830^{+0.019}_{-0.014} \end{bmatrix}$	$\begin{bmatrix} 0.805^{+0.014}_{-0.013} & -0.010^{+0.002}_{-0.002} \\ -0.026^{+0.003}_{-0.002} & 0.180^{+0.003}_{-0.002} \end{bmatrix}$	$\begin{bmatrix} -0.211^{+0.004}_{-0.005} & 0.201^{+0.006}_{-0.005} \\ 0.241^{+0.006}_{-0.005} & 0.820^{+0.020}_{-0.015} \end{bmatrix}$
AGN mask weight=0	$\begin{bmatrix} 0.736^{+0.017}_{-0.016} & -0.291^{+0.006}_{-0.007} \\ -0.307^{+0.007}_{-0.007} & 0.332^{+0.008}_{-0.008} \end{bmatrix}$	$\begin{bmatrix} -0.103^{+0.004}_{-0.004} & 0.255^{+0.007}_{-0.007} \\ 0.286^{+0.008}_{-0.008} & 0.868^{+0.022}_{-0.021} \end{bmatrix}$	$\begin{bmatrix} 0.837^{+0.019}_{-0.018} & -0.007^{+0.002}_{-0.002} \\ -0.023^{+0.003}_{-0.003} & 0.185^{+0.005}_{-0.005} \end{bmatrix}$	$\begin{bmatrix} -0.219^{+0.007}_{-0.008} & 0.211^{+0.007}_{-0.007} \\ 0.250^{+0.008}_{-0.007} & 0.863^{+0.024}_{-0.023} \end{bmatrix}$
AGN mask+1pix	$\begin{bmatrix} 0.737^{+0.016}_{-0.015} & -0.290^{+0.006}_{-0.006} \\ -0.306^{+0.007}_{-0.007} & 0.334^{+0.008}_{-0.007} \end{bmatrix}$	$\begin{bmatrix} -0.106^{+0.004}_{-0.004} & 0.256^{+0.006}_{-0.006} \\ 0.286^{+0.008}_{-0.007} & 0.868^{+0.021}_{-0.020} \end{bmatrix}$	$\begin{bmatrix} 0.840^{+0.017}_{-0.017} & -0.008^{+0.002}_{-0.002} \\ -0.023^{+0.003}_{-0.003} & 0.187^{+0.005}_{-0.004} \end{bmatrix}$	$\begin{bmatrix} -0.223^{+0.007}_{-0.007} & 0.216^{+0.006}_{-0.006} \\ 0.253^{+0.007}_{-0.007} & 0.863^{+0.023}_{-0.022} \end{bmatrix}$
AGN mask+2pix	$\begin{bmatrix} 0.734^{+0.016}_{-0.015} & -0.290^{+0.005}_{-0.006} \\ -0.306^{+0.006}_{-0.007} & 0.334^{+0.008}_{-0.007} \end{bmatrix}$	$\begin{bmatrix} -0.105^{+0.004}_{-0.004} & 0.256^{+0.006}_{-0.006} \\ 0.285^{+0.007}_{-0.008} & 0.863^{+0.022}_{-0.019} \end{bmatrix}$	$\begin{bmatrix} 0.837^{+0.018}_{-0.016} & -0.008^{+0.002}_{-0.002} \\ -0.023^{+0.003}_{-0.003} & 0.186^{+0.005}_{-0.005} \end{bmatrix}$	$\begin{bmatrix} -0.223^{+0.007}_{-0.007} & 0.216^{+0.006}_{-0.006} \\ 0.253^{+0.008}_{-0.006} & 0.858^{+0.023}_{-0.020} \end{bmatrix}$
5 perturbers	$\begin{bmatrix} 0.707^{+0.012}_{-0.011} & -0.285^{+0.004}_{-0.004} \\ -0.301^{+0.005}_{-0.005} & 0.317^{+0.006}_{-0.005} \end{bmatrix}$	$\begin{bmatrix} -0.090^{+0.003}_{-0.003} & 0.249^{+0.006}_{-0.004} \\ 0.277^{+0.006}_{-0.005} & 0.826^{+0.015}_{-0.012} \end{bmatrix}$	$\begin{bmatrix} 0.802^{+0.014}_{-0.012} & -0.006^{+0.002}_{-0.002} \\ -0.019^{+0.002}_{-0.002} & 0.181^{+0.004}_{-0.003} \end{bmatrix}$	$\begin{bmatrix} -0.216^{+0.005}_{-0.006} & 0.202^{+0.005}_{-0.005} \\ 0.237^{+0.006}_{-0.005} & 0.841^{+0.017}_{-0.013} \end{bmatrix}$
Sérsic profiles	$\begin{bmatrix} 0.688^{+0.019}_{-0.015} & -0.271^{+0.006}_{-0.007} \\ -0.285^{+0.007}_{-0.008} & 0.309^{+0.009}_{-0.008} \end{bmatrix}$	$\begin{bmatrix} -0.098^{+0.004}_{-0.004} & 0.236^{+0.007}_{-0.006} \\ 0.262^{+0.008}_{-0.007} & 0.801^{+0.025}_{-0.019} \end{bmatrix}$	$\begin{bmatrix} 0.784^{+0.022}_{-0.018} & -0.008^{+0.002}_{-0.002} \\ -0.022^{+0.002}_{-0.003} & 0.173^{+0.005}_{-0.004} \end{bmatrix}$	$\begin{bmatrix} -0.205^{+0.007}_{-0.008} & 0.198^{+0.008}_{-0.007} \\ 0.231^{+0.008}_{-0.007} & 0.790^{+0.027}_{-0.021} \end{bmatrix}$
Composite	$\begin{bmatrix} 0.737^{+0.034}_{-0.026} & -0.281^{+0.009}_{-0.010} \\ -0.295^{+0.010}_{-0.011} & 0.315^{+0.012}_{-0.010} \end{bmatrix}$	$\begin{bmatrix} -0.104^{+0.006}_{-0.007} & 0.251^{+0.009}_{-0.008} \\ 0.276^{+0.011}_{-0.010} & 0.837^{+0.028}_{-0.025} \end{bmatrix}$	$\begin{bmatrix} 0.833^{+0.028}_{-0.027} & -0.016^{+0.007}_{-0.015} \\ -0.029^{+0.006}_{-0.016} & 0.180^{+0.007}_{-0.006} \end{bmatrix}$	$\begin{bmatrix} -0.209^{+0.008}_{-0.010} & 0.212^{+0.008}_{-0.008} \\ 0.244^{+0.010}_{-0.009} & 0.834^{+0.029}_{-0.026} \end{bmatrix}$
Composite,AGN mask+1pix	$\begin{bmatrix} 0.738^{+0.025}_{-0.026} & -0.291^{+0.011}_{-0.010} \\ -0.304^{+0.012}_{-0.011} & 0.329^{+0.011}_{-0.012} \end{bmatrix}$	$\begin{bmatrix} -0.103^{+0.006}_{-0.006} & 0.257^{+0.009}_{-0.010} \\ 0.282^{+0.011}_{-0.013} & 0.855^{+0.028}_{-0.030} \end{bmatrix}$	$\begin{bmatrix} 0.843^{+0.027}_{-0.028} & -0.008^{+0.002}_{-0.002} \\ -0.022^{+0.003}_{-0.003} & 0.183^{+0.007}_{-0.007} \end{bmatrix}$	$\begin{bmatrix} -0.217^{+0.011}_{-0.010} & 0.221^{+0.009}_{-0.008} \\ 0.253^{+0.011}_{-0.011} & 0.855^{+0.030}_{-0.032} \end{bmatrix}$
Composite,Arcmask+1pix,50x50src	$\begin{bmatrix} 0.705^{+0.012}_{-0.013} & -0.282^{+0.005}_{-0.005} \\ -0.298^{+0.006}_{-0.006} & 0.319^{+0.006}_{-0.006} \end{bmatrix}$	$\begin{bmatrix} -0.097^{+0.003}_{-0.003} & 0.243^{+0.004}_{-0.005} \\ 0.275^{+0.006}_{-0.006} & 0.814^{+0.013}_{-0.014} \end{bmatrix}$	$\begin{bmatrix} 0.795^{+0.013}_{-0.014} & -0.008^{+0.001}_{-0.001} \\ -0.025^{+0.002}_{-0.002} & 0.178^{+0.003}_{-0.004} \end{bmatrix}$	$\begin{bmatrix} -0.208^{+0.005}_{-0.005} & 0.196^{+0.004}_{-0.005} \\ 0.236^{+0.005}_{-0.006} & 0.811^{+0.014}_{-0.015} \end{bmatrix}$

4.2.5 A single H_0 from multiple lenses

With the current facilities at our disposal to work on time-delay cosmography, the best-case scenario for individual lens systems yields a time-delay distance at $\sim 5\%$ precision. Although that uncertainty could be ideally reduced to roughly 2-3% through the use of next-generation instruments (JWST, E-ELT, ...) combined to high-cadence monitoring campaigns, reaching a $\sim 1\%$ precision on H_0 requires to combine the measurements from multiple lenses. That may look like a very simple and natural thing to do, yet we must ensure that we are *allowed* to do it.

In theory, two measurements carried out on two different lenses should yield consistent values of H_0 (or any other parameters from the underlying cosmological model), providing that i) the measurements have been carried out correctly, without omitting any known or unknown sources of bias and ii) the cosmological models in which the comparison takes place predict that two values measured on different lenses must be the same. Beyond the eventuality of a statistical fluke, any tension between two or more measurements must be interpreted as a sign that one of the two conditions above is not met. To know if we can combine various lenses, we follow the Bayesian formalism proposed in Marshall et al. (2006). We consider the hypothesis $\mathrm{H}^{\mathrm{global}}$ expressing that the time-delay distance distributions from various lenses can be simultaneously fit using a common set of cosmological parameters. In opposition, the hypothesis $\mathrm{H}^{\mathrm{ind}}$ assumes that at least one lens would be better represented using another, independent set of cosmological parameters. To compare the likeliness of these two hypothesis, we make use of the Bayes factor F that makes the $\mathrm{H}^{\mathrm{global}}$ hypothesis F times more likely. The Bayes factor is computed as follows:

$$F = \frac{P(\mathbf{d_1},\mathbf{d_2},\mathbf{d_3}|\mathrm{H}^{\mathrm{global}})}{P(\mathbf{d_1}|\mathrm{H}^{\mathrm{ind}})\mathrm{P}(\mathbf{d_2}|\mathrm{H}^{\mathrm{ind}})\mathrm{P}(\mathbf{d_3}|\mathrm{H}^{\mathrm{ind}})}, \tag{4.8}$$

where d_i represents a given data set, in our case the time-delay distance likelihood from a lens. To be consistent with our hypothesis definition, we must compute the Bayes factor F i) using the three lenses simultaneously, ii) using individual pairs of lenses and iii) combining each pair with the third lens. By denoting the time-delay distance likelihood of the lens i by $\mathbf{L_i}$ and its mean by squared brackets, the three cases above put in Eq. 4.8 reduce to the following (see the Appendix of Suyu et al., 2013, for details)

$$F_{1\cup 2} = \frac{\langle \mathbf{L_1 L_2} \rangle}{\langle \mathbf{L_1} \rangle \langle \mathbf{L_2} \rangle}, \quad F_{1\cup 2\cup 3} = \frac{\langle \mathbf{L_1 L_2 L_3} \rangle}{\langle \mathbf{L_1} \rangle \langle \mathbf{L_2} \rangle \langle \mathbf{L_3} \rangle}, \quad F_{12\cup 3} = \frac{\langle \mathbf{L_1 L_2 L_3} \rangle}{\langle \mathbf{L_1 L_2} \rangle \langle \mathbf{L_3} \rangle} \tag{4.9}$$

A Bayes factor F bigger than 1 means that the $\mathrm{H}^{\mathrm{global}}$ is favored by the data over $\mathrm{H}^{\mathrm{ind}}$. Although arbitrary, we can define a scale to interpret the value of F. Following for example Jeffreys (1998), the evidence of $\mathrm{H}^{\mathrm{global}}$ against $\mathrm{H}^{\mathrm{ind}}$ becomes *substantial* for $F > 5$, *strong* for $F > 10$

and *very strong* for $F > 15$. However, it is worth noting that even a very strong evidence in favor of $\mathrm{H}^{\mathrm{global}}$ does not mean at all that $\mathrm{H}^{\mathrm{ind}}$ must be rejected ! If we take, for comparison, two Gaussian likelihoods then $F = 1$ when the two functions overlap at their 2σ points, and is of $F \sim 3.6$ when they overlap at their 1σ points. What we must ensure here is that F is significantly larger than 1, meaning that a common cosmological model fits all the data sets well enough so that they can thus be combined. For the three H0LiCOW lenses, the lowest value obtained is $F = 1.1$ when combining in an open-ΛCDM Universe RXJ1131-1231 with B1608+656. All the other combinations in all the considered cosmological models yields a higher Bayes factor. For example, combining the three lenses simultaneously in a flat-ΛCDM Universe yields $F = 14.2$. In that case, it is (very) roughly equivalent to say the data speak at $\sim 94.5\%$ in favor of a common cosmological model, from which we conclude that we can combine the three lenses without any significant loss of information. The combination is thus simply made by importance sampling the individual lenses on each other.

As mentioned earlier, when cosmological models others than ΛCDM are considered, the definition of the comoving distance (Eq. 4.7) has to be adapted. The weak dependency of D_{Δ_t} on the cosmological parameters other than H_0 results in no drastic changes of the inferred H_0 values. Yet, it is still interesting to compute the H_0 distribution in various models, at least to compare and combine it with other experiments. Through our analysis, we consider four cosmological models in which we compute the cosmological constraints from time-delay cosmography alone. First, the UH_0 model that corresponds to a flat-ΛCDM assumes a fixed value of $\Omega_m = 0.32$, following the most recent measurements from (Planck Collaboration et al., 2016c). Then, the flat- and open-ΛCDM, respectively abridged UΛCDM and UoΛCDM that have uniform priors on the relevant cosmological parameters. Finally, the UwCDM that replace the cosmological constant ΛCDM by a form of dark energy whose equation of state is written as $w = p/\rho$ and where w is a free parameter. The models are detailed in Tab. 2 of (Bonvin et al., 2017) and partially reproduce in Fig. 4.6, that presents the individual and combined H_0 distribution in these four cosmological models.

The H_0 values in the UH_0, UΛCDM and UoΛCDM are all consistent with each other. The UwCDM is higher but has larger error bars; the sample distribution plotted in the two-dimensional plane H_0 versus w (see Fig. 5 of Bonvin et al., 2017) indicates that higher values of H_0 are correlated with lower values of w, whose prior was purposely loosely constraining. In a flat-ΛCDM Universe (second panel from the top), we find a value of $H_0 = 71.9^{+2.4}_{-3.0}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, i.e. at 3.8% precision. By reducing the prior range on Ω_m to [0, 0.5], we found a slightly higher value of $H_0 = 72.5^{+2.4}_{-2.4}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, at 3.3% precision, much closer to the UH_0 value. The former value was chosen in (Bonvin et al., 2017) as the final estimate from our sample of three lenses, as plotted in Fig. 4.1. However, we can reasonably argue that reducing the Ω_m prior range to [0, 0.5] should not limit our analysis, since it is fairly well admitted that matter cannot fill up the whole content of the Universe and reproduce the current observations, unless gravity is modified at large scale in a very, very peculiar way.

To quantify the discordance between the various measurements of H_0 in a given cosmological model, we can compute the *tension* between them. If two independent experiments yield two results of the same measurement $A^{+a_+}_{-a_-}$ and $B^{+b_+}_{-b_-}$ with $A > B$, we define the tension in σ units as $(A-B)/\sqrt{b_+^2 + a_-^2}$. Two Gaussian distributions overlapping at their 1σ points have a 1σ tension, and the same distributions overlapping at their 2σ points a tension of $\sim 2.8\sigma$. Let us now consider the revised value from time-delay cosmography of $H_0 = 72.5^{+2.4}_{-2.4}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, the distance ladder measurement from Riess et al. (2016) of $H_0 = 73.24 \pm 1.74\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and the up-to-date intermediate results from Planck that corrects from systematic effects in the polarization maps ($H_0 = 66.93 \pm 0.62\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ Planck Collaboration et al., 2016d). With these numbers, we obtain a tension of 2.2σ between strong lenses and Planck, 0.2σ between strong lenses and distance ladder and 3.4σ between distance ladder and Planck. When combining the strong lenses and distance ladder into a single local estimate of $H_0^{\rm local} = 72.93 \pm 1.43$, we find a 3.8σ tension between the local and Planck values.

4.2.6 Cosmology beyond ΛCDM

Correctly interpreting the tension between two measurements is a very delicate thing to do. Roughly speaking, a 2σ tension between two Gaussian distributions could arise due to a purely statistical fluke with a 5% chance. At 3σ, the chance is only of 0.3%. Yet, as previously expressed, measurements and associated error bars could be strongly affected by unknown unknowns, notwithstanding the efforts deployed to avoid them. Furthermore, the distributions used to compute the tension are not necessarily Gaussian, which should be taken into account; if Riess et al. (2016) claims a 3.4σ tension with Planck Collaboration et al. (2016d), a reanalysis of the two data sets in a fully Bayesian framework by Feeney et al. (2017) similar to the Bayes factor analysis presented in Sec 4.2.5 conclude that the odds against ΛCDM are at best substantial (7 against 1), at worst strong (10 against 1)[4]. Thus, despite the apparent tension between the local and CMB measurements of H_0 in a flat-ΛCDM universe, the situation is certainly less dramatic than it sounds when speaking about the tension considering the σ's only. It would then be overconfident to claim that the flat-ΛCDM model is at fault. Nevertheless, the tension speaks in favor of exploring other cosmological models beyond ΛCDM, notably to see if there is a way to alleviate the disagreement. In such cases, the data sets from various experiments can be combined to infer more precise values of the cosmological parameters.

For example, in Bonvin et al. (2017), we explore four *1-parameter extensions* to ΛCDM. We consider in turn an open universe in the oΛCDM model, an universe with dark energy characterized by a free[5] equation of state $w = p/\rho$ in the wCDM model, an universe with free effective number of relativistic neutrino species $\mathrm{N_{eff}}$ in the $\mathrm{N_{eff}}\Lambda$CDM model and an universe with a free total mass of neutrinos $\Sigma \mathrm{m}_\nu$ in the $\mathrm{m}_\nu\Lambda$CDM model. Our priors on these models corresponds to the CMB posterior distributions (either WMAP or Planck), on which we importance sample our results. It is interesting to note that in these models, the CMB results are much

---

[4] The use of "best" and "worst" adjective here is completely subjective. Trying not to hurt any sensitivities, I chose the conservative way of thinking.

[5] Here, *free* means that the value of the considered parameter is not fixed *a priori*.

less constraining than in a flat-ΛCDM model, due to the degeneracies arising in the global fit. The local probes of H_0 being nearly independent of other cosmological parameters, they are thus of great help in constraining cosmological models, as already stated in Linder (2011); Weinberg et al. (2013). Figure 4.7 presents the constraints from multiple probes combined altogether in the four 1-parameter extensions named above.

By giving an extra degree of freedom to the CMB experiments and (sometimes) to time-delay cosmography as well, the individually inferred values of H_0 are shifted from their flat-ΛCDM counterpart, with a precision that generally worsen. As a result, the tension lowers to $\sim 2\sigma$ and less. With that tension reduced and thus the data set agreeing reasonably well with each other, the interesting thing to do is to combine them and see how it constrains the extra parameter of the considered cosmological model. In the oΛCDM model, the curvature density from Planck and strong lensing reads $\Omega_k = 0.003^{+0.004}_{-0.006}$, compatible with a flat universe. In the m$_\nu\Lambda$CDM model, the upper bound for the total mass of neutrino is constrained at $\Sigma \mathrm{m}_\nu \leq 0.182$ eV at 95% confidence. Both constraints can be tightened further by adding BAO, CMB lensing and Supernovae as extra constraints. This illustrates well the *complementarity* of these different probes.

In the N$_{\mathrm{eff}}\Lambda$CDM model, the predicted effective number of relativistic neutrino species by Planck and time-delay cosmography is $\mathrm{N}_{\mathrm{eff}} = 3.45^{+0.23}_{-0.24}$, i.e. $\sim 1.7\sigma$ above the $\mathrm{N}_{\mathrm{eff}} = 3.046$ value assumed in the standard model. In the wCDM model, the predicted value from Planck and time-delay cosmography for the dark energy equation of state parameter reads $w = -1.38^{+0.14}_{-0.16}$, more than 2σ below the value for a cosmological constant $w = -1$ advocated in the ΛCDM model. It must be noted that this last prediction is in a $\sim 2\sigma$ tension with Planck and BAO combined, that are compatible with a dark energy in the form of a cosmological constant.

In summary, the complementarity of Planck and time-delay cosmography in these ΛCDM extensions allows both to alleviate the original tension and to explore the possibility of physics beyond the standard cosmological model, here in the form of extra relativistic species of neutrinos ($\mathrm{N}_{\mathrm{eff}} > 3.046$) or phantom dark energy ($w < -1$). However, the fact that other complementary data sets such as BAO are still not in agreement with time-delay cosmography in these cosmologies beyond ΛCDM should at least prevent us from drawing any definitive conclusions. One could argue that the solution might be to go further and consider 2 - or even more - parameters extensions to ΛCDM (for example, Di Valentino et al. (2017) consider a 12-parameters extension), but adding more and more degrees of freedom is opposed to the Occam's razor principle, and there is so far no compelling evidence that such extra models are needed; in fact, although tensions arise in the ΛCDM as demonstrated with the value of H_0, the evidence is not yet strong enough to rightfully convince the scientific community to move away from it. It has to be noted, however, that recent measurements of the cosmic shear by KiDS (Hildebrandt et al., 2017; Köhlinger et al., 2017) (although not everyone seems to agree, see e.g. van Uitert et al. (2017); Efstathiou and Lemos (2017)) or dark matter profile at the center of bright cluster galaxies (Harvey et al., 2017) are also reportedly in tension with the ΛCDM model. Cosmology now stands at a crossroad: are these tensions real so that models beyond flat-ΛCDM should be seriously considered (e.g. Bernal et al., 2016; Di Valentino et al., 2016; Karwal and Kamionkowski, 2016; Di Valentino et al., 2017) or do they result from nasty statistical flukes and/or unknown unknowns that do not hold against more complete analysis (e.g. Cardona et al., 2017; Heavens et al., 2017; Simpson et al., 2017)? Although the data seem to speak in favor of the former option, only more observational evidence will tell us with certainty which path should be followed.

Bonvin et al. (2017), reproduced below, presents the details of the time-delay measurements of HE0435-1223 addressed in chapters 2 and 3, and also details the cosmological analysis from the three H0LiCOW lenses addressed in the present chapter. This paper, along with the other papers of the H0LiCOW series, has been advertised in a press release detailed in Appendix B.

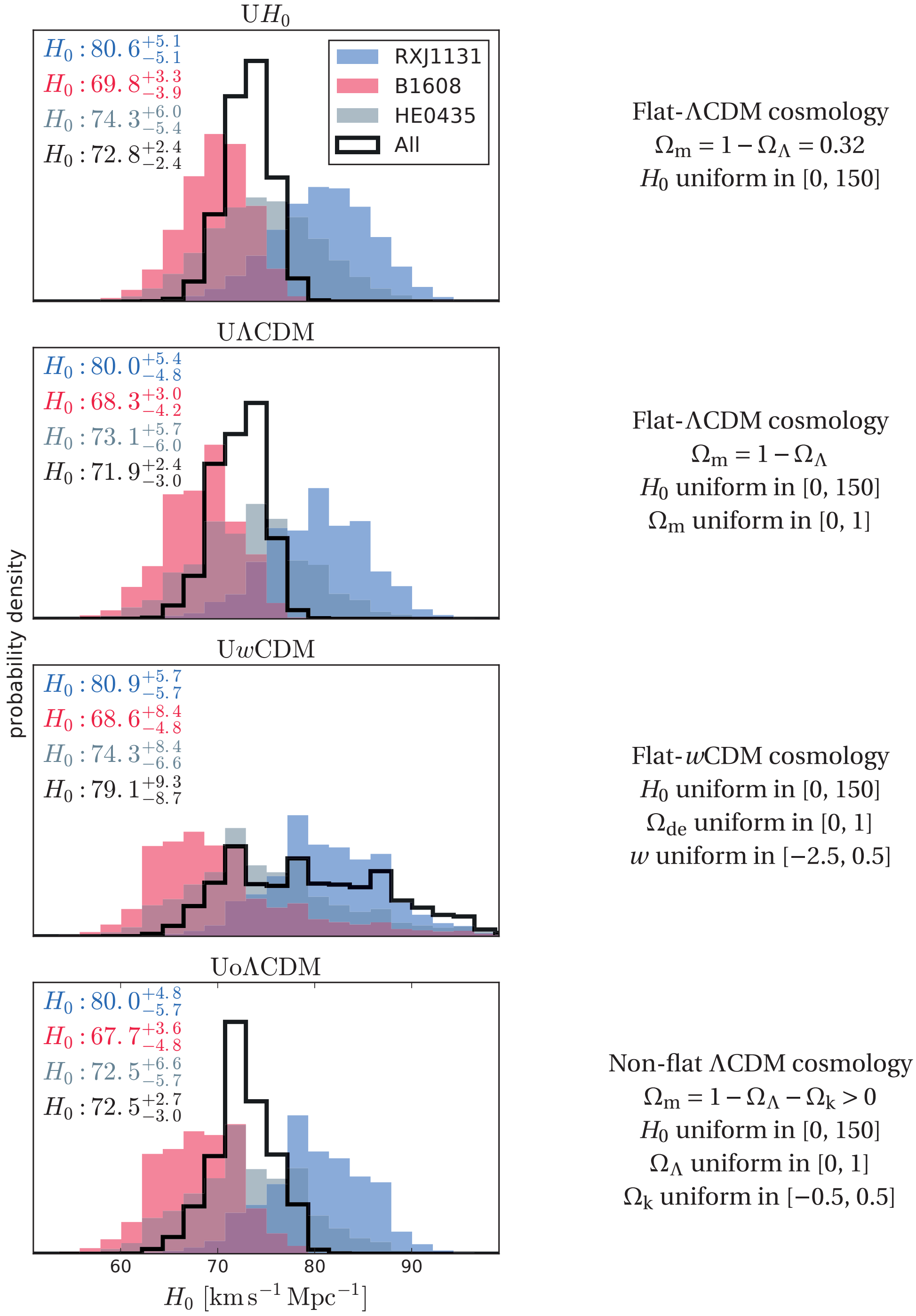


Figure 4.6: Marginalized posterior probability distributions for H_0 in the UH_0, UΛCDM, UwCDM and UoΛCDM cosmologies using the constraints from B1608+656, RXJ1131-1231 and HE0435-1223. The colored histograms present the distributions for each individual lens, whereas the solid black line corresponds to the combination of the three data sets. The quoted values of H_0 in the top-left corner of each panel are the median, 16th and 84th percentiles. The priors used on each model are indicated on the right column. Adapted from Bonvin et al. (2017).

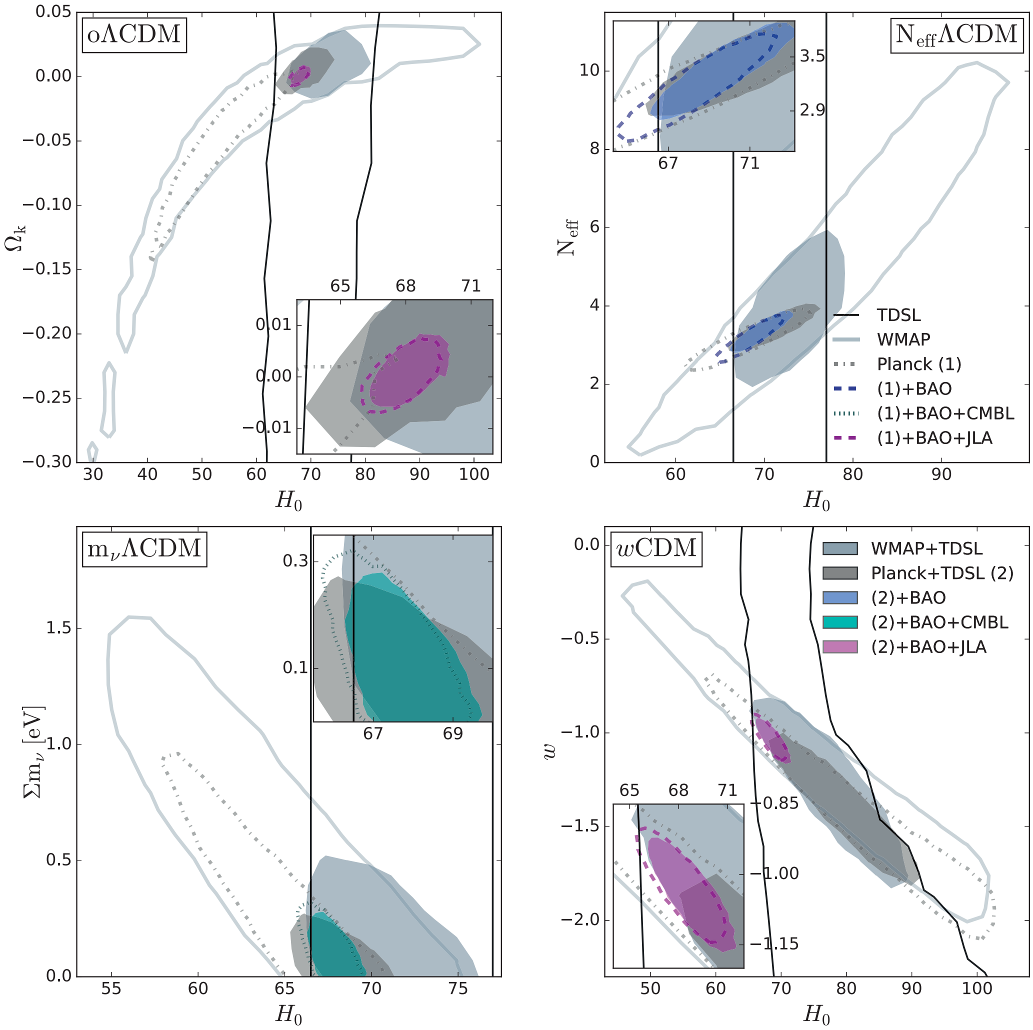


Figure 4.7: Cosmological constraints from various experiments in 1-parameter extensions to ΛCDM. We consider a non-zero curvature parameter Ω_k in the top-left panel, a variable effective number of relativistic neutrino species N$_{\rm eff}$ in the top-right panel, a variable total mass of neutrinos Σm$_\nu$ in the bottom-left panel and a time-invariant dark energy equation of state $w = p/\rho$ in the bottom-right panel. The filled regions and colored lines delimit the marginalized 95% confidence interval, with and without applying the constraints from time-delay cosmography, respectively. The different colors represent the constraints drawn from WMAP or from Planck, possibly combined with time-delay cosmography (TDSL), CMB lensing from Planck (CMBL), Baryon Acoustic Oscillations (BAO) and a large sample of Supernovae (JLA) - see Planck Collaboration et al. (2016c) for details. The solid black lines delimit the 95% confidence region for strong lensing alone in the corresponding uniform cosmology with no additional information. Adapted from Bonvin et al. (2017).

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **465,** 4914–4930 (2017) doi:10.1093/mnras/stw3006
Advance Access publication 2016 November 22

H0LiCOW – V. New COSMOGRAIL time delays of HE 0435−1223: H_0 to 3.8 per cent precision from strong lensing in a flat ΛCDM model

V. Bonvin,[1★] F. Courbin,[1] S. H. Suyu,[2,3,4] P. J. Marshall,[5] C. E. Rusu,[6] D. Sluse,[7]
M. Tewes,[8] K. C. Wong,[9,4] T. Collett,[10] C. D. Fassnacht,[7] T. Treu,[11]
M. W. Auger,[12] S. Hilbert,[13,14] L. V. E. Koopmans,[15] G. Meylan,[1]
N. Rumbaugh,[11] A. Sonnenfeld[16,11,17] and C. Spiniello[2]

Affiliations are listed at the end of the paper

Accepted 2016 November 17. Received 2016 November 14; in original form 2016 July 5

ABSTRACT

We present a new measurement of the Hubble Constant H_0 and other cosmological parameters based on the joint analysis of three multiply imaged quasar systems with measured gravitational time delays. First, we measure the time delay of HE 0435−1223 from 13-yr light curves obtained as part of the COSMOGRAIL project. Companion papers detail the modelling of the main deflectors and line-of-sight effects, and how these data are combined to determine the time-delay distance of HE 0435−1223. Crucially, the measurements are carried out blindly with respect to cosmological parameters in order to avoid confirmation bias. We then combine the time-delay distance of HE 0435−1223 with previous measurements from systems B1608+656 and RXJ1131−1231 to create a Time Delay Strong Lensing probe (TDSL). In flat Λ cold dark matter (ΛCDM) with free matter and energy density, we find $H_0 = 71.9^{+2.4}_{-3.0}\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$ and $\Omega_\Lambda = 0.62^{+0.24}_{-0.35}$. This measurement is completely independent of, and in agreement with, the local distance ladder measurements of H_0. We explore more general cosmological models combining TDSL with other probes, illustrating its power to break degeneracies inherent to other methods. The joint constraints from TDSL and Planck are $H_0 = 69.2^{+1.4}_{-2.2}\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_\Lambda = 0.70^{+0.01}_{-0.01}$ and $\Omega_\mathrm{k} = 0.003^{+0.004}_{-0.006}$ in open ΛCDM and $H_0 = 79.0^{+4.4}_{-4.2}\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_\mathrm{de} = 0.77^{+0.02}_{-0.03}$ and $w = -1.38^{+0.14}_{-0.16}$ in flat wCDM. In combination with Planck and baryon acoustic oscillation data, when relaxing the constraints on the numbers of relativistic species we find $\mathrm{N_{eff}} = 3.34^{+0.21}_{-0.21}$ in $\mathrm{N_{eff}}$ΛCDM and when relaxing the total mass of neutrinos we find $\Sigma m_\nu \leq 0.182$ eV in m_νΛCDM. Finally, in an open wCDM in combination with Planck and cosmic microwave background lensing, we find $H_0 = 77.9^{+5.0}_{-4.2}\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_\mathrm{de} = 0.77^{+0.03}_{-0.03}$, $\Omega_\mathrm{k} = -0.003^{+0.004}_{-0.004}$ and $w = -1.37^{+0.18}_{-0.23}$.

Key words: gravitational lensing: strong – galaxies: individual: HE 0435−1223 – cosmology: observations – distance scale.

1 INTRODUCTION

In the past decade, the Standard Cosmological Model, Λ cold dark matter (ΛCDM), which assumes the existence of either a cosmological constant or a form of dark energy with equation of state $w = -1$, and large-scale structure predominantly composed of cold dark matter, has been firmly established given observations to date (e.g. Hinshaw et al. 2013; Planck Collaboration XIII 2016a). From a minimal set of six parameters describing ΛCDM, one can in principle infer the value of other parameters such as the current expansion rate of the Universe, H_0. However, such an inference involves strong assumptions about the cosmological model, such as the absence of curvature or a constant equation of state for the dark energy. Conversely, we can relax these assumptions and explore models beyond flat ΛCDM using a wider set of cosmological probes. In this case, the analysis benefits greatly from independent measurements of H_0 from observations of distance probes such as the distance ladder or water masers (see e.g. Treu 2010; Weinberg et al. 2013; Treu & Marshall 2016, for a review). As Weinberg et al. (2013) point out, the figure of merit of any stage III or stage IV cosmological survey improves by 40 per cent if an independent measurement of H_0 is available to a precision of 1 per cent.

The ‘time-delay distances’ in gravitationally lensed quasar systems offer an opportunity to measure H_0 independently of any

---

★E-mail: vivien.bonvin@epfl.ch.

Published by Oxford University Press on behalf of the Royal Astronomical Society

other cosmological probe. First suggested by Refsdal (1964), this approach involves measuring the time delays between multiple images of a distant source produced by a foreground lensing object. The time delays depend on the matter distribution in the lens (galaxy), on the overall matter distribution along the line of sight and on the cosmological parameters. The time delays are related to the so-called time-delay distance $D_{\Delta t}$ to the lens and the source, which is primarily sensitive to H_0 and has a weak dependence on the matter density $\Omega_{\rm m}$, the dark energy density $\Omega_{\rm de}$, the dark energy equation of state, w, and on the curvature parameter Ω_k (e.g. Suyu et al. 2010; Linder 2011).

The first critical step for the method to work is the measurement of the time delays from a photometric monitoring campaign to measure the shift in time between the light curves of the lensed images of quasars. Such monitoring campaigns must be long enough, and have good enough temporal sampling, to catch all possible (and usually small) photometric variations in the light curves. This is the goal of the COSMOGRAIL collaboration: the COSmological MOnitoring of GRAvItational Lenses, which has been monitoring about 20 lensed quasars with 1-m class and 2-m class telescopes since 2004 (e.g. Courbin et al. 2005; Eigenbrod et al. 2006a; Bonvin et al. 2016). The target precision for the time-delay measurements is a few per cent or better, because the error on the time delays propagates linearly to the first order on the cosmological distance measurement. Examples of COSMOGRAIL results include Courbin et al. (2011), Tewes et al. (2013b), Rathna Kumar et al. (2013) and Eulaers et al. (2013).

The second critical step is the modelling of the lens galaxy. Indeed, time-delay measurements alone can constrain only a combination of the time-delay distance and the surface density of the lens around the quasar images (Kochanek 2002). Additional constraints on the density profile of the lens are therefore required in order to convert observed time delays into inferences of the time-delay distance. These constraints can be derived from velocity dispersion measurements, and the radial magnification of the extended, lensed arc image of the quasar host galaxy (e.g. Suyu et al. 2010, 2014). Ideal targets for this purpose are lensed quasars with a prominent host, which offer strong constraints on the density profile slope of the foreground lens.

In modelling the lens mass distribution, special care has to be paid to the mass-sheet degeneracy (MSD), and, more generally, the source-position transformation (SPT; e.g. Falco, Gorenstein & Shapiro 1985; Wucknitz 2002; Schneider & Sluse 2013, 2014; Unruh, Schneider & Sluse 2016; Xu et al. 2016). These can be seen as degeneracies in the choice of the gravitational lensing potential that leave all the lensing observables invariant except for the modelled time delay, Δt. In other words, a wrong model of the main lens mass distribution can perfectly fit the observed morphology of the lensing system, and yet result in an inaccurate inference of the time-delay distance. Priors and spectroscopic constraints on the dynamics of the main lens therefore play a critical role in avoiding systematic biases. In addition, perturbations to the lens potential by the distribution of mass along the line of sight also creates degeneracies in the lens modelling. The latter can be mitigated with a measurement of the mass distribution along the line of sight, for example by using spectroscopic redshift measurements of the galaxies in the lens environment (e.g. Fassnacht et al. 2006; Wong et al. 2011), comparisons between galaxy number counts in the real data and in simulations (Hilbert et al. 2007, 2009; Fassnacht, Koopmans & Wong 2011; Collett et al. 2013; Greene et al. 2013; Suyu et al. 2013; McCully et al. 2016) or using weak-lensing measurements (Tihhonova et al., in preparation)

The H0LiCOW collaboration (H_0 Lenses in COSMOGRAIL's Wellspring) capitalizes on the efforts of COSMOGRAIL to measure accurate time delays, and on high-quality auxiliary data from *Hubble Space Telescope* (*HST*) and 10-m class ground-based telescopes, to constrain cosmology. The H0LiCOW sample consists of five well-selected targets, each with exquisite time-delay measurements. B1608+656, monitored in radio band with the VLA (Fassnacht et al. 2002), and RXJ1131−1231, monitored by COSMOGRAIL in the visible (Tewes et al. 2013b), have already shown promising results, with relative precisions on the time-delay distance of 5 and 6.6 per cent, respectively (Suyu et al. 2010, 2014).

This paper is part of the H0LiCOW series, focusing on the quadruple lensed quasar HE 0435−1223 (α(2000): $04^{\rm h}38^{\rm m}14\overset{\rm s}{.}9$; δ(2000): $-12^\circ17'14\overset{''}{.}4$) (Wisotzki et al. 2000, 2002) discovered during the Hamburg/ESO Survey (HES) for bright quasars in the Southern hemisphere. The source redshift has been measured by Sluse et al. (2012) as $z_{\rm s} = 1.693$, and the redshift of the lens has been measured by Morgan et al. (2005) and Eigenbrod et al. (2006b) as $z_{\rm d} = 0.4546 \pm 0.0002$. The lens lies in a group of galaxies of at least 12 members. A first measurement of the time delay for HE 0435−1223 was presented in Courbin et al. (2011). In this work, we present a significant improvement of the time-delay measurement, with twice as long light curves as in Courbin et al. (2011). The other H0LiCOW papers include an overview of the project (Suyu et al., submitted; hereafter H0LiCOW Paper I), a spectroscopic survey of the field of HE 0435−1223 and a characterization of the groups along the line of sight (Sluse et al., submitted; hereafter H0LiCOW Paper II), a photometric survey of the field of HE 0435−1223 with an estimate of the effect of the external line-of-sight structure (Rusu et al., submitted; hereafter H0LiCOW Paper III), and a detailed modelling of the lens and the inference of the time-delay distance along with cosmological results for HE 0435−1223 (Wong et al., in press; hereafter H0LiCOW Paper IV). In this paper, we combine the results for HE 0435−1223 with those from the other two lensed quasars already published, and with other cosmological data sets (Bennett et al. 2013; Hinshaw et al. 2013; Planck Collaboration XIII 2016a).

This paper is organized as follows. Section 2 presents the COSMOGRAIL optical monitoring data and its reduction process. Section 3 presents the time-delay measurements and related uncertainties. Section 4 summarizes the main steps of the field-of-view analysis detailed in H0LiCOW Paper II and H0LiCOW Paper III and the lens modelling detailed in H0LiCOW Paper IV that lead to the time-delay distance determination. Section 5 combines the time-delay distance of HE 0435−1223 and other lenses, and with additional cosmological data sets, in order to make the best possible inferences of cosmological parameters. Finally, Section 6 presents our conclusions and future prospects in the light of these results.

2 PHOTOMETRIC MONITORING DATA

HE 0435−1223 has been monitored since 2003 as part of the COSMOGRAIL programme and in collaboration with the Kochanek et al. (2006) team. The data acquired from autumn 2003 to spring 2010 were presented in Courbin et al. (2011). Here, we double the monitoring period, adding observations taken between autumn 2010 and spring 2016. Our monitoring sites include two Northern telescopes: the 1.2 m Belgian Mercator telescope located at the Roque de Los Muchachos Observatory, La Palma, Canary Islands (Spain) and the 1.5 m telescope located at the Maidanak Observatory (Uzbekistan). The average observing cadence was 11 and 16 d, respectively, at these sites. These telescopes ceased taking data for

Table 1. Optical monitoring campaigns of HE 0435−1223. The sampling is the mean number of days between the observations, not considering the seasonal gaps.

Telescope	Camera	FoV	Pixel	Period of observation	#obs	Exp. time	Median FWHM	Sampling
Euler	C2	11 arcmin × 11 arcmin	0.344 arcsec	2004 Jan–2010 Mar	301	5 × 360 s	1.37 arcsec	6 d
Euler	ECAM	14.2 arcmin × 14.2 arcmin	0.215 arcsec	2010 Sep–2016 Mar	301	5 × 360 s	1.39 arcsec	4 d
Mercator	MEROPE	6.5 arcmin × 6.5 arcmin	0.190 arcsec	2004 Sep–2008 Dec	104	5 × 360 s	1.59 arcsec	11 d
Maidanak	SITE	8.9 arcmin × 3.5 arcmin	0.266 arcsec	2004 Oct–2006 Jul	26	10 × 180 s	1.31 arcsec	16 d
Maidanak	SI	18.1 arcmin × 18.1 arcmin	0.266 arcsec	2006 Aug–2007 Jan	8	6 × 300 s	1.31 arcsec	16 d
SMARTS	ANDICAM	10 arcmin × 10 arcmin	0.300 arcsec	2003 Aug–2005 Apr	136	3 × 300 s	≤1.80 arcsec	4 d
TOTAL	–	–	–	2003 Aug–2016 Mar	876	394.5 h	–	3.6 d

COSMOGRAIL in 2008 December. In the Southern hemisphere, the Swiss 1.2 m Euler telescope located at the ESO La Silla observatory (Chile) has monitored HE 0435−1223 since 2004. Two cameras were used: the C2 and the EulerCAM instruments, with an average cadence of 6 and 4 d, respectively. We also make use of the data obtained at the 1.3 m SMARTS ANDICAM camera at Cerro Tololo Inter-American Observatory. Note that we do not re-analyse the SMARTS data, but use directly the published photometric measurements (Kochanek et al. 2006). Table 1 gives a detailed summary of the observations.

2.1 Data reduction

The full data set consists of two distinct blocks that do not overlap in time and that we treat independently. The first block includes the Mercator, Maidanak and Euler-C2 data, to which we add the published SMARTS photometry. The detailed processing and the relative photometric calibration of these curves is presented in section 2.2 of Courbin et al. (2011). The second block consists of the 301 new data points obtained with EulerCAM that we reduce with the pipeline described in section 3 of Tewes et al. (2013b), whose main steps can be summarized as follows.

(i) Each image is corrected for bias and readout effects. We then apply a flat-field correction using a high signal-to-noise master sky-flat which we correct for a pattern generated by the shutter opening and closing times. A spatially variable sky background frame is then constructed using the SExtractor software (Bertin & Arnouts 1996) and we subtract it from the data frame. All the frames are aligned and analysed to carry out the photometric measurements. Fig. 1 presents a stack of the 100 EulerCAM images with a seeing smaller than 1.14 arcsec.

(ii) The photometric measurements of the four blended images of HE 0435−1223 are obtained using deconvolution photometry using the MCS deconvolution algorithm (Magain, Courbin & Sohy 1998; Cantale et al. 2016). To do this, the point spread function (PSF) is measured, for each exposure individually, using the seven stars labelled PSF1–PSF7 on Fig. 1. A simultaneous deconvolution of all the frames is then carried out, leading to a model composed of a deep image representing extended sources, and a catalogue of point sources with improved resolution and sampling. During the deconvolution process, the data are decomposed into a sum of analytical point sources (the quasar images) and of a numerical pixel channel containing the image of the lensing galaxy and of any potential extended object.

(iii) We compute a multiplicative median normalization coefficient for each exposure, using several deconvolved reference stars. If possible at all, we use stars whose colours is similar to that of the quasar. In the case of HE 0435−1223, we use eight reference stars, labelled N1–N8 in Fig. 1. We then apply the normalization coefficient to the deconvolved images of the point sources. Their intensities are returned for every frame, hence leading to the light curves.

The upper panel of Fig. 2 presents the 13-yr-long COSMOGRAIL light curves of HE 0435−1223, including the data from Courbin et al. (2011) and our new data. The similarity between the four light curves is immediately noticeable. However, it can also be noted that they would not superpose perfectly when shifted in time and magnitude, due to ‘extrinsic variability’ which is interpreted as being caused by microlensing by stars in the lensing galaxy (see e.g. Blackburne et al. 2014; Braibant et al. 2014). These extrinsic contributions are clearly seen here on time-scales from a few weeks to several years, in the form of an evolution of the magnitude-separation between the light curves. They must be handled properly in order to measure time delays with high accuracy.

2.2 On the importance of long light curves

Given the limited photometric precision of the COSMOGRAIL images, long-term monitoring is crucial to the time-delay measurement for two main reasons. First, one needs to catch enough intrinsic photometric variations in the quasar light curves in order to identify common structures. In the present case, these can be found on average two to three times per observing season, with some seasons displaying more prominent structures than others. Inflexion points in the light curves are most precious to constrain the time delays. For example, dips and peaks with an amplitude of nearly half a magnitude can be observed in several seasons: 2004–2005, 2012–2013 and 2015–2016. Secondly, the extrinsic variability related to microlensing must be taken into account (e.g. by modelling and removing it) to avoid time-delay measurement biases. Any simple and well-constrainable model is likely not sufficient to capture all aspects of this extrinsic variability, and might result in residual biases. The availability of decade long light curves allows us to check for potentially significant biases by analysing subsets of the full data, and certainly to reduce residual ones.

3 TIME-DELAY MEASUREMENT

With the light curves in hand, the time delays can be measured using numerical techniques accounting for noisy photometry, irregular temporal sampling and seasonal monitoring gaps. These techniques must also account for the extrinsic variability in the quasar images, related to microlensing effects, to avoid systematic error on the time-delay measurements. Different techniques have been devised in the literature to carry out this task, and the COSMOGRAIL collaboration has implemented its own approach and several algorithms (see Tewes, Courbin & Meylan 2013a, also for a summary of extrinsic variability causes). These techniques are

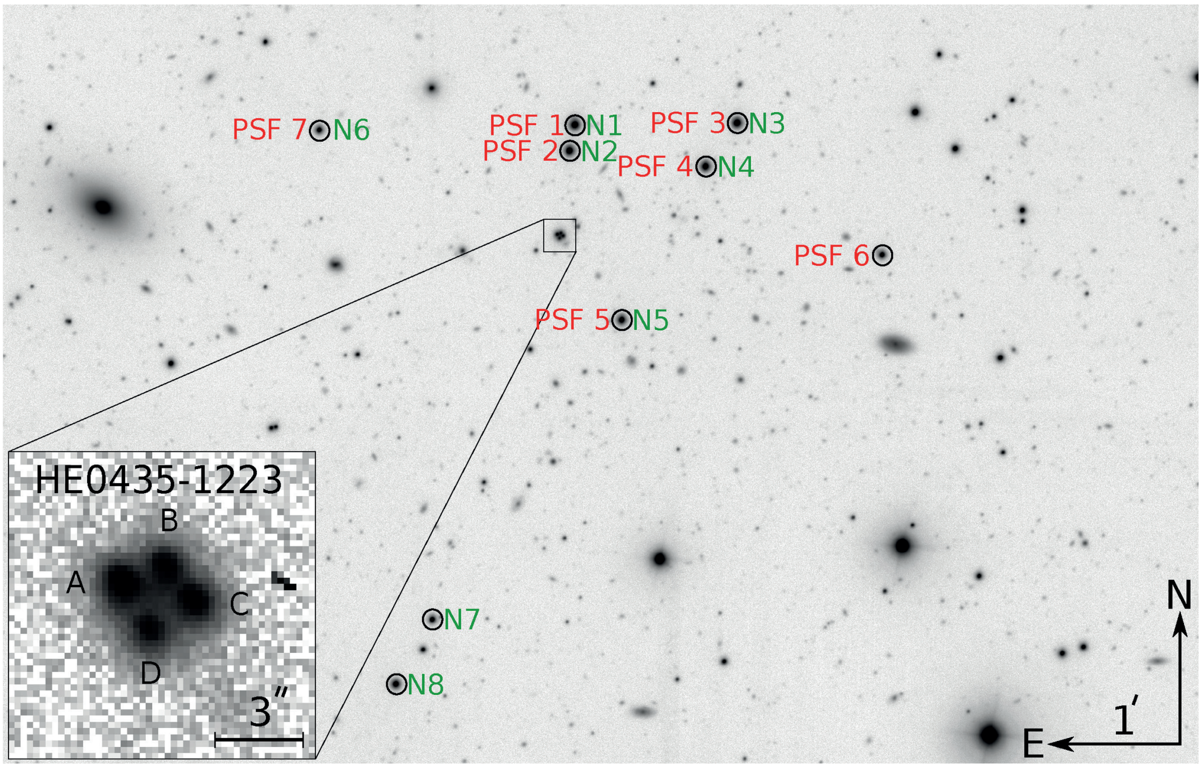


Figure 1. Part of the field of view of EulerCAM installed on the Swiss 1.2 m telescope around the quasar HE 0435−1223. This image is a combination of 100 exposures of 360 s each, for a total exposure time of 10 h. The stars used to build a PSF model for each EulerCAM exposure are circled and labelled PSF1 to PSF7 in red, and the stars used for the photometric calibrations are circled and labelled N1–N8 in green. The insert in the bottom left shows the single, 360 s exposure of the lens, for reference. Note that photometric and spectroscopic redshifts are available for many galaxies in the field of view (see H0LiCOW Paper II and H0LiCOW Paper III for details).

publicly available as a PYTHON package named PYCS.[1] They have been tested using realistic numerical simulations, and have been confronted with the data provided to the lensing community by the first Time-Delay Challenge (TDC1; see Dobler et al. 2015; Liao et al. 2015). An in-depth analysis of their performance proved them to be both precise and accurate (Bonvin et al. 2016) under various observational conditions, and in particular for light curves mimicking the COSMOGRAIL data. Among the three point-estimation algorithms provided in the PYCS toolbox, we consider for this work two algorithms based on very different principles.

(i) *The free-knot spline technique* models the light curves as a sum of intrinsic variations of the quasar, common to the four light curves, plus some extrinsic variability different in each of the four light curves. The algorithm simultaneously fits one continuous curve for the intrinsic variations, four less-flexible curves for the extrinsic variations, and time shifts between the four light curves. All curves are represented as *free-knot* splines (see e.g. Molinari, Durand & Sabatier 2004), for which the knot locations are optimized at the same time as the spline coefficients and the time shifts.

(ii) *The regression difference technique* minimizes the variability of the difference between Gaussian-process regressions performed on each light curve. This method has no explicitly parametrized model for extrinsic variability. Instead, it yields time-delay estimates which minimize apparent extrinsic variability on time-scales comparable to that of the precious intrinsic variability features. We see the contrasting approaches of this technique and the free-knot splines as valuable to detect potential method-related biases, and will use the regression difference technique as a cross-check of our results in this paper.

The third original PYCS estimator, a dispersion technique that was inspired by Pelt et al. (1996) and used in the previous analysis of HE 0435−1223 (Courbin et al. 2011) has proven to be less accurate in several investigations of simulated data (see Eulaers et al. 2013; Rathna Kumar et al. 2013; Tewes et al. 2013a,b). For this reason, we do not consider it in this work.

We stress that the uncertainty estimation for the time delays is at least as important as the above point estimators. It is carried out within PYCS by assessing the point-estimation performance on synthetic light curves. This approach attempts to capture significantly more than the formal uncertainty which could be derived from the photometric error bars, if one would assume that for instance the spline model described above is a sufficient description of the data.

3.1 Application to the data

To apply the free-knot spline and regression difference techniques provided by PYCS to our data, we closely follow the procedure

---

[1] PYCS can be obtained from http://www.cosmograil.org

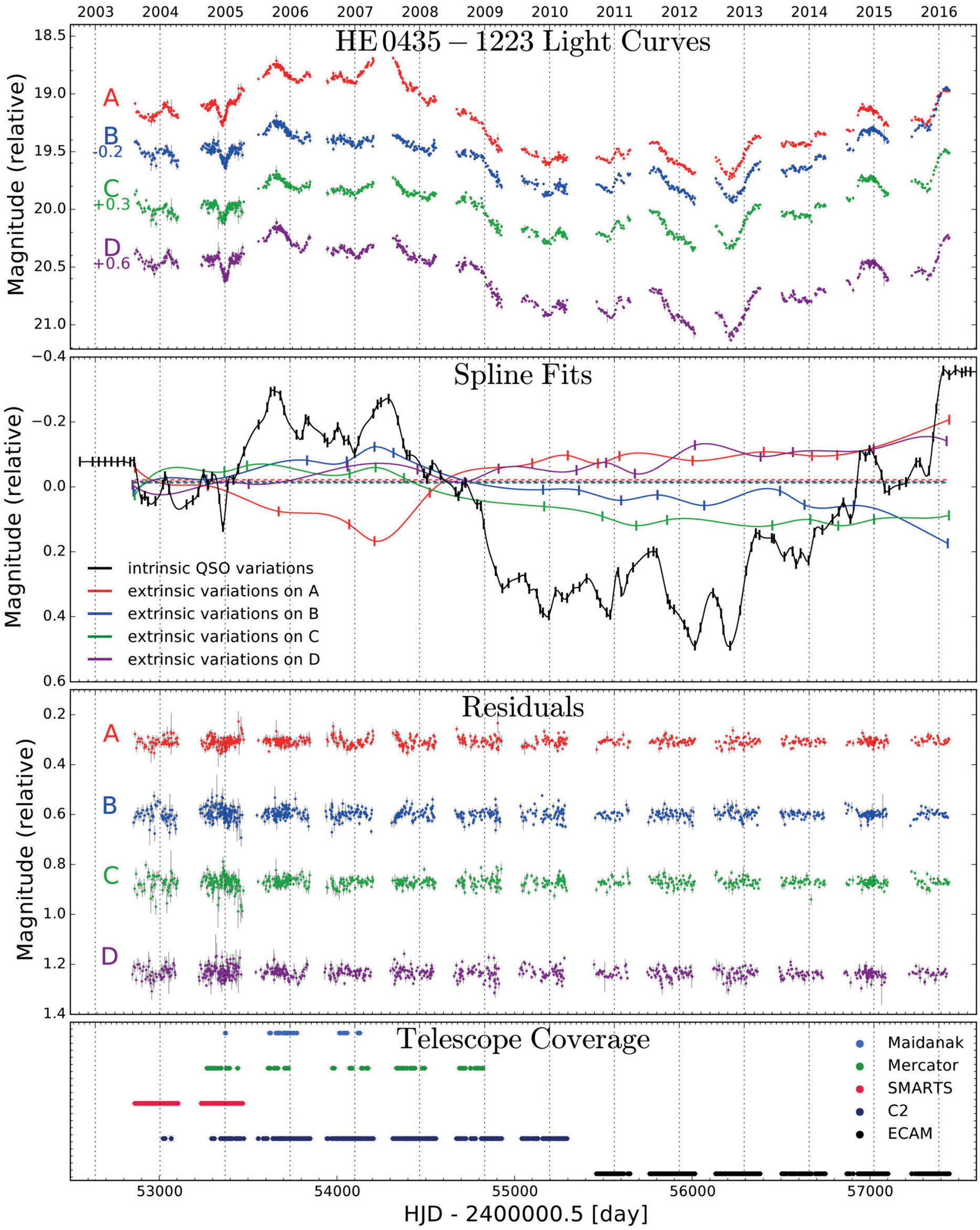


Figure 2. From top to bottom: light curves for the four lensed images of the quasar HE 0435−1223. The relative shifts in magnitude are chosen to ease visualization, and do not influence the time-delay measurements. The second panel shows a model of the intrinsic variations of the quasar (black) and the four curves for the extrinsic variations in each quasar image using the free-knot spline technique (colour code). The vertical ticks indicate the position of the spline knots. The residuals of the fits for each light curve is shown in the next panel. Finally, the bottom panel displays the journal of the observations for HE 0435−1223 for the five telescopes or cameras used to gather the data over 13 yr (see column '#obs' of Table 1), where each point represents one monitoring night. The light curves are publicly available on the H0LiCOW,[2] COSMOGRAIL[3] and CDS websites.

---

[2] H0LiCOW : www.h0licow.org

[3] COSMOGRAIL : www.cosmograil.org

described in Tewes et al. (2013a), and summarized in the following.[4] A key ingredient of this approach is the careful generation of mock light curves which are used to fine-tune and assess the precision and bias of the point estimators. These simulations are fully synthetic, in the sense that they are drawn from models with known time delays (hereafter *true* time delays), and yet they closely mimic the quasar variability signal and the extrinsic variability from the observed data. The PYCS free-knot spline technique is used to create the generative models from which we draw these simulations. For this, we start by fitting an intrinsic spline with on average 10 knots yr^{-1} and four extrinsic splines with 2 knots yr^{-1} to the observations. These average knot densities are sufficiently high to fit all unambiguous patterns observed in the data, while still resulting in a negligible intrinsic variance, i.e. avoiding significant degeneracies between the time-delay estimates and the spline models. Such a free-knot spline fit is illustrated in the second panel of Fig. 2. The extrinsic variability splines presented here, when subtracted to each other pair-wise, are compatible with the data presented in Blackburne et al. (2014). However, we also show in our robustness checks [see point (ii) of Section 3.2] that variations in the modelling of the extrinsic variability do not influence much the time-delay measurements.

Before drawing the synthetic mock curves by sampling from this model fit, the smooth extrinsic splines are locally augmented with fast correlated noise. This noise follows a power-law spectrum which is iteratively adjusted so that the scatter in the mock curves has similar statistical properties to the scatter measured in the observed data. We then draw 1000 mock data sets, with true time shifts uniformly distributed within ±3 d around our best-fitting solution. This results in a range of ±6 d for the true delays, largely covering all plausible situations for this lens system. It is important to use simulations with various true time delays to tune and/or verify the accuracy of the point estimators. Tests on simulations with only a single true time delay would not probe bias and precision reliably, especially as many time-delay estimators are prone to responding unsteadily to the true delay.

The third panel of Fig. 2 shows the observed residual light curves after subtraction of a free-knot spline fit, and the bottom panel depicts the coverage by the different telescopes and instruments. During the first 5 yr of monitoring, three to four different telescopes were used, with a mean residual dispersion of all data points of $\sigma = 25$ mmag. During the last years (2011 to present) one telescope was used, with a mean residual dispersion of $\sigma = 15$ mmag. Besides unmodelled microlensing effects, part of this scatter comes from night-to-night and instrument-to-instrument calibration of the data. Long-term monitoring programmes of gravitational lenses are a matter of balance between the gain in temporal sampling using multiple telescopes, and the losses in photometric precision due to combining data from different instruments. Future monitoring programmes will need to account for this trade-off (Courbin et al., in preparation).

We run the free-knot spline fit and the regression difference technique (with a Matérn covariance function, an amplitude parameter of 2.0 mag, a scale of 250 d and a smoothness degree $\nu = 1.5$) on the observed light curves as well as on the mocks (for details, see Tewes et al. 2013a). Fig. 3 presents our time-delay estimates along with their 1σ uncertainties, and compares them to the previous measurements by Courbin et al. (2011), for which the dispersion technique was used. The uncertainties are computed by summing the maximum estimated bias and statistical uncertainty in quadrature. The free-knot spline technique and regression difference technique are in relatively good agreement with each other, with a maximum tension of 1.3σ. Recall that the measurements are not independent, and therefore good agreement is to be expected. The two techniques also yield a similar precision, with a 6.5 per cent relative uncertainty on the longest delay, i.e. Δt_{AD}.

3.2 Robustness checks

In order to test the robustness of our time-delay measurements, we performed several simple checks.

(i) We carried out several times the deconvolution of the ECAM data, using PSF stars and/or normalization stars that differ from the ones adopted in Fig. 1. We also changed the initial parameters of the MCS deconvolution photometry. These include an estimate for the light profile of the lens galaxy, the astrometry of the quasar images and of the lens galaxy and the flux of the quasar images at each epoch. All these changes resulted in a slightly higher scatter in the ECAM light-curves data points, yet without significant impact on the time-delay measurements.

(ii) We varied the intrinsic and/or extrinsic variability model of the free-knot spline technique by changing the number of knots used. We used 8–12 knots yr^{-1} for the intrinsic model, and 0.5–2 knots yr^{-1} for the extrinsic model. Free-knot splines have the advantage over regular splines or polynomials that their ability to fit prominent variability features is less sensitive to the total number of parameters. Using a lower or higher number of knots did not significantly affect the time-delay measurements. The residual light curves (third panel of Fig. 2) remain statistically similar.

(iii) Taking advantage of the 13 yr of monitoring, we split the light curves into three parts: (i) seasons 2003–2004 to 2006–2007, (ii) seasons 2007–2008 to 2011–2012 and (iii) seasons 2012–2013 to 2015–2016. We measured the time delays independently on each of these subsections. The results are presented in the bottom parts of each panel of Fig. 3. We see that the measurements on these subsections are well distributed around the delays measured on the full light curves. Furthermore, a clear majority of the delays obtained on the subsections cover, within the given 1σ error estimates, the results from the full curves. To conclude, these robustness checks give no strong evidence that the achieved time-delay uncertainties are significantly underestimated and/or biased.

3.3 Time delays of HE 0435−1223

We have shown that our two curve shifting techniques lead to comparable time delays and error bars on the full light curves of HE 0435−1223, which is reassuring. Still, one needs to define which time-delay estimates to propagate into the time-delay distance (H0LiCOW Paper IV) and cosmological parameter inferences. We opt for using the results from the free-knot spline technique. This method has been tested extensively on a broad range of simulated light curves and proved to be both precise and accurate (Bonvin et al. 2016). In addition, Sluse & Tewes (2014) showed with this same technique that a flexible extrinsic variability model can prevent potential time-delay biases due to the delayed emission of the broad-line region of the quasar with respect to the accretion disc.

4 TIME-DELAY DISTANCE

The time delays determined in Section 3, combined with a careful modelling of the lens galaxy mass distribution, can be used to infer

---

[4] For the sake of reproducibility, the complete PYTHON code used to measure the delays is available at http://www.cosmograil.org.

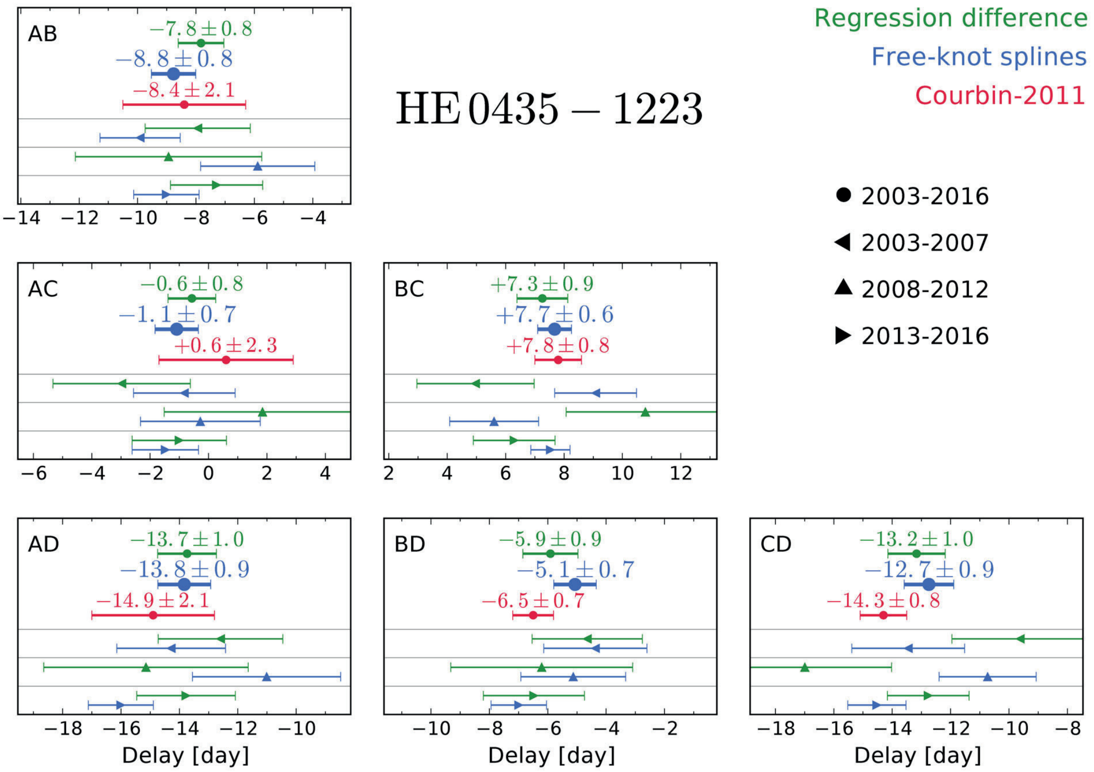


Figure 3. Time delays for the six pairs of quasar images, as indicated in top-left corner of each panel. In each panel, we show the time-delay measurement along with the 1σ error bar using our two best curve-shifting techniques, and compare with the measurement of Courbin et al. (2011). We also show the result of measurements carried out with the free-knot spline technique and regression difference technique when splitting the data in three continuous chunks of 4 or 5 yr each. All cosmological results in this work use the time-delay measurements from the free-knot splines (larger blue symbols on the figure).

the time-delay distance in the HE 0435−1223 system. The lens modelling and time-delay distance determination are addressed in detail in H0LiCOW Paper IV and are only summarized here.

4.1 Principles of the measurement

The time delay Δt_{ij} between two lensed images of the same object can be written as follows:

$$\Delta t_{ij} = \frac{D_{\Delta t}}{c}\left[\frac{(\boldsymbol{\theta}_i - \boldsymbol{\beta})^2}{2} - \psi(\boldsymbol{\theta}_i) - \frac{(\boldsymbol{\theta}_j - \boldsymbol{\beta})^2}{2} + \psi(\boldsymbol{\theta}_j)\right], \quad (1)$$

where $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$ are the coordinates of the images i and j in the lens plane, $\boldsymbol{\theta}$ is the position of the lensed images on the plane of the sky, $\boldsymbol{\beta}$ is the unlensed source position and $\psi(\boldsymbol{\theta}_i)$ is the lens potential at position $\boldsymbol{\theta}_i$. The time-delay distance $D_{\Delta t}$ is defined to be the following combination of three angular diameter distances and the deflector (i.e. the lens) redshift z_d: $D_{\Delta t} \equiv (1 + z_\mathrm{d})D_\mathrm{d}D_\mathrm{s}/D_\mathrm{ds}$. Here, D_d, D_s and D_ds are, respectively, the angular distances between the observer and the deflector, the observer and the source and the deflector and the source. The time-delay distance is, by construction, proportional to the inverse of the Hubble constant H_0^{-1}, and is primarily sensitive to this of all cosmological parameters. A posterior probability distribution for $D_{\Delta t}$ allows us to infer a probability distribution for H_0, assuming a given cosmology.

In the case of HE 0435−1223, there are multiple galaxies at different redshifts close in projection to the strong lens system. We explicitly include these galaxies in our multilens plane lens model in H0LiCOW Paper IV, and in doing so introduce more angular diameter distances into the problem. However, we can still form the posterior predictive distribution for the 'effective' time-delay distance defined above, and it is the latter that we use to infer cosmological parameters. All the remaining additional mass along the line of sight can also weakly focus and defocus the light rays from the source, an effect that needs to be corrected for. We model this external contribution using an external convergence term κ_ext that modifies the time-delay distance as follows:

$$D_{\Delta t} = \frac{D_{\Delta t}^\mathrm{model}}{1 - \kappa_\mathrm{ext}}. \quad (2)$$

Here, $D_{\Delta t}^\mathrm{model}$ is the effective time-delay distance predicted by the multiplane model, and $D_{\Delta t}$ is the corrected time-delay distance we seek. Given probability density functions (PDFs) for $P(D_{\Delta t}^\mathrm{model})$ and κ_ext, we can compute the PDF for $D_{\Delta t}$. In H0LiCOW Paper IV, we derive a lognormal approximation for $P(D_{\Delta t})$, and it is this that we use as a likelihood function $P(D_{\Delta t}|\boldsymbol{\theta}, H)$ for cosmological parameters $\boldsymbol{\theta}$ given a cosmological model H.

In the rest of this section, we provide a brief summary of each part of the analysis just outlined, before proceeding to the cosmological parameter inference in Section 5.

4.2 Determination of the external convergence

We use two complementary approaches to quantify the impact of the mass along the line of sight, both yielding consistent results.

4.2.1 Spectroscopy of the field

In H0LiCOW Paper II, we perform a spectroscopic identification of a large fraction of the brightest galaxies[5] located within a projected distance of 3 arcmin of the lens. This catalogue is complemented with spectroscopic data from Momcheva et al. (2015) that augment redshift measurements to projected distances of ∼15 arcmin from the lens. Based on those data, we show that, from the five galaxies located within 12 arcsec of the lens, the galaxy G1 ($z = 0.782$), closest in projection, produces the largest perturbation of the gravitational potential, and hence needs to be explicitly included in the lens models. The other galaxies are found to produce significantly smaller perturbations. On the other hand, we search for galaxy groups and clusters that would be massive enough to modify the structure of the lens potential, but find none. On the lower mass end (i.e. groups with $\sigma \leq 500$ km s^{-1}), nine group candidates are found in the vicinity of the lens. We demonstrate that none of the groups discovered are massive enough/close enough in projection to produce high-order perturbation of the gravitational lens potential (McCully et al. 2014, 2016). This is also confirmed by a weak-lensing analysis of the field of HE 0435−1223 (Tihhonova et al., in preparation)

4.2.2 Weighted galaxy number counts

In H0LiCOW Paper III, we calculate the probability distribution for the external convergence using a weighted galaxy number counts technique (Greene et al. 2013). We conduct a wide-field, broad-band optical to mid-infrared photometric survey of the field in order to separate galaxies from stars, determine the spatial distribution of galaxies around HE 0435−1223, and estimate photometric redshifts and stellar masses. We compare weighted galaxy number counts around the lens, given an aperture and flux limit, to those through similar apertures and flux limits in CFHTLenS (Heymans et al. 2012). We investigate weights that incorporate the projected distance and redshift to the lens as well as the galaxy stellar masses. The resulting number under/overdensities serve as constraints in selecting similar fields from the Millennium Simulation, and their associated $\kappa_{\rm ext}$ values, from the catalogue of Hilbert et al. (2009). We find that the resulting distribution of $\kappa_{\rm ext}$ is consistent with the typical mean density value (i.e. $\kappa_{\rm ext} = 0$) and is robust to choices of weights, apertures, flux limits and cosmology, up to an impact of 0.5 per cent on the time-delay distance.

4.3 Mass modelling

In H0LiCOW Paper IV, we perform our lens modelling using GLEE, a software package developed by A. Halkola and S. H. Suyu (Suyu & Halkola 2010; Suyu et al. 2012). Our fiducial mass model for the lens galaxy is a singular power-law elliptical mass distribution with external shear. We explicitly include the closest line of sight perturbing galaxy in the lens model (G1; see fig. 3 of H0LiCOW Paper IV), using the full multiplane lens equation to account for its effects. We also include in an extended modelling four other nearby perturbing galaxies to check their impact. Because the perturbers are at different redshifts, there is no single time-delay distance that can be clearly defined. Instead, we vary H_0 directly in our models and then use this distribution to calculate an effective time-delay distance, where the angular diameter distances $D_{\rm d}$ and $D_{\rm s}$ are calculated using the redshift of the main deflector, $z_{\rm d} = 0.454$. We assume a fiducial cosmology, $\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$ and $w = -1$ in this modelling procedure, but we find that allowing these cosmological parameters to vary has a negligible (<1 per cent) effect on the resulting effective time-delay distance distribution.

The MSD – the invariance to the lensed images under addition of a uniform mass sheet to our mass model combined with a rescaling of the source plane coordinates – can affect the inferred time delays, and may limit the effectiveness of time delays in constraining cosmology (e.g. Schneider & Sluse 2013, 2014). We have shown in previous work that including the central velocity dispersion of the main galaxy in the lens modelling minimizes the effect of the MSD (see fig. 4 of Suyu et al. 2014). In the case of HE 0435−1223, we measure $\sigma = 222 \pm 15\,{\rm km\,s^{-1}}$ using Keck I spectroscopy. We also show that mass models that go beyond the elliptically-symmetric power-law profile, and that are better physically justified, fit our data equally well yet lead to the same cosmological inference. As in Suyu et al. (2014), the H0LiCOW Paper IV tests both power law and a composite model with a baryonic component and an NFW dark matter halo. We also note that the completely independent models of Birrer, Amara & Refregier (2016) confirm the findings of Suyu et al. (2014).

We model the images of the lensed source simultaneously in three *HST* bands: ACS/$F555W$, ACS/$F814W$ and WFC3/$F160W$. The lensed quasar images are modelled as point sources convolved with the PSF. The extended, unlensed image of the host galaxy of the quasar is modelled separately on a pixel grid with curvature regularization (see e.g. Suyu et al. 2006). Our constraints on the model include the positions of the quasar images, the measured time delays and the surface brightness pixels in each of the three bands. Model parameters of the lens are explored through Markov Chain Monte Carlo (MCMC) sampling, while the Gaussian posterior PDF for the source pixel values is characterized using standard linear algebra techniques (e.g. Suyu et al. 2006).

During our modelling procedure, we iteratively update the PSFs using the lensed AGN images themselves in a manner similar to Chen et al. (2016), and we use these corrected PSFs in our final models (for more details, see Suyu et al., in preparation). We conduct multiple robustness tests to account for various systematic uncertainties in the modelling. We vary our choice of modelling regions and use various weights for each pixel. We use various assumptions for the light profiles fits of the lens galaxy and we model the lens using alternative mass models, comparing the use of power-law profiles and chameleon profiles. We also explicitly include the five nearest perturbing galaxies into our modelling. All the models are given a similar weight, reflecting the possible choices available through the analysis, and are combined together to yield a single posterior PDF for $D_{\Delta t}$. Fig. 9 of H0LiCOW Paper IV presents the individual and combined posterior distributions, highlighting on one hand the relatively good agreement between the models, and on the other hand the need to consider a sufficiently flexible model to fully take into account as many sources of systematics as possible.

4.4 Blinding methodology and unblinded results

A key element of our analysis is that it is carried out blindly with respect to the inference of cosmological parameters. This blindness is crucial in order to avoid unconscious confirmation bias. In practice, blindness is built into our measurement in the following manner. All

---

[5] The completeness of the spectroscopic identification depends on the distance to the lens and limiting magnitude, see fig. 3 of H0LiCOW Paper II. For example, 60 per cent (80 per cent) of the galaxies brighter than $i \sim 22$ mag ($i \sim 21.5$ mag) have a measured spectroscopic redshift within a radius of 3 arcmin (2 arcmin) of the lens.

the individual measurements and modelling efforts in H0LiCOW Paper IV are carried out without any knowledge of the effects of specific choices on the resulting cosmology. In some cases, this blindness is trivial to achieve; for example, the measurement of velocity dispersion was carried out and finalized independently from the cosmological inference, and the connection between the two is significantly complex and indirect that the person carrying out the velocity dispersion measurement effectively had no way to determine how that could affect cosmological parameters. In other cases, building on the procedure established by our previous analysis of RXJ1131−1231 (Suyu et al. 2013), blindness was achieved by only using plotting codes that offset every posterior probability distribution for time-delay distance and cosmological parameters by a constant (such as the median value of each marginal distribution), and thus never revealing the actual measurements to the investigators until the time of unblinding (see discussion in H0LiCOW Paper IV).

All of our analysis and visualization tools were developed and tested using simulated quantities. No modifications were allowed after the official unblinding, making the unblinding step irreversible. The official unblinding was originally scheduled for 2016 June 2 during a teleconference open to all the co-authors. Additional tests were suggested during this meeting. As a result, the analysis was kept blind for another two weeks and the final unblinding happened during a teleconference starting at 6 AM UT on 2016 June 16 and was audio recorded by LVEK without others knowing until the end of the teleconference. The results presented in the next section are the combination of the blind measurements obtained for HE 0435−1223 and RXJ1131−1231,[6] and the not-blind measurements obtained by our team for the first system B1608+656.

5 JOINT COSMOGRAPHY ANALYSIS

Cosmic microwave background (CMB) experiments provide a model-dependent value of the Hubble constant, H_0, which appear to be in some tension with methods based on standard rulers and standard candles. In a flat-ΛCDM universe, the significance of the tension between the most recent values from Planck (Planck Collaboration XIII 2016a) and the direct measurement from Cepheids and Type Ia Supernovae (Riess et al. 2016) is 3.3σ. Either this tension is due, at least in part, to systematics in the measurements (as suggested by e.g. Efstathiou 2014), or it is caused by new physics beyond the predictions of flat ΛCDM. Several authors discuss the possibility of relaxing the usual assumptions about cosmological parameters as a way to reduce the tension (e.g. Salvatelli et al. 2013; Heavens, Jimenez & Verde 2014; Di Valentino, Melchiorri & Silk 2016). Possible assumptions include, for example, that we live in a non-flat universe ($\Omega_k \neq 0$), that the dark energy equation of state is not a cosmological constant ($w \neq -1$), that the sum of the neutrino masses is larger than predicted by the standard hierarchy scenario ($\Sigma m_\nu > 0.06$ eV), and/or that the effective number of relativistic neutrino species may differ from its assumed value in the standard model ($N_{\rm eff} \neq 3.046$). Given the above, it is important to consider a range of plausible extended cosmological models when investigating the information that can be gained from any specific cosmological probe (see e.g. Collett & Auger 2014; Giusarma et al. 2016).

In this context, we present below our inference of the cosmological parameters obtained using the time-delay distance measurements of the strongly lensed quasars B1608+656, WFI2033−4723 and HE 0435−1223. After making sure that their individual results are consistent with each other, we present our cosmological inference using all three systems jointly, referred as 'TDSL' for 'Time Delay Strong Lensing'. We then combine TDSL with the *WMAP* Data Release 9 (hereafter '*WMAP*'; Bennett et al. 2013; Hinshaw et al. 2013) and with the Planck 2015 Data Release[7] (Planck Collaboration XIII 2016a, hereafter 'Planck'). When available, we also use the combination of Planck data with Planck measurements of CMB weak-lensing (hereafter 'CMBL'; Planck Collaboration XV 2016b), with baryon acoustic oscillation surveys at various redshifts (hereafter BAO; Percival et al. 2010; Beutler et al. 2011; Blake et al. 2011; Anderson et al. 2012; Padmanabhan et al. 2012) and with the data of the Joint Lightcurve Analysis of Supernovae (hereafter 'JLA'; Betoule et al. 2013). The latter two data sets are described in detail in section 5.2 of Planck Collaboration XVI et al. (2014) and section 5.3 of Planck Collaboration XIII (2016a), respectively. Note that when possible, we do not combine the cosmological probes other than TDSL ourselves, but instead we use the combined results published and provided by the Planck team.[8]

We follow the importance sampling approach suggested by Lewis & Bridle (2002) and employed by Suyu et al. (2010, 2013), reweighting the *WMAP* and Planck posterior samples with the TDSL likelihoods from the analyses of B1608+656 (Suyu et al. 2010), RXJ1131−1231 (Suyu et al. 2014) and HE 0435−1223 (H0LiCOW Paper IV), for the set cosmological models described in Table 2.

We consider both (i) 'uniform' cosmologies, with only a few variable cosmological parameters with uniform priors in order to get constraints from TDSL alone, and (ii) cosmologies extended beyond ΛCDM where we combine the TDSL likelihoods with other probes. When comparing two cosmological parameter inferences, we use the following terminology. When two results differ by less than 1σ, we consider that they are 'consistent'; in the 1σ–2σ range they are in 'mild tension'; in the 2σ–3σ range they are in 'tension'; above 3σ, they are in 'significant tension'. If a cosmological parameter inference follows a non-Gaussian distribution, we use '1σ' to refer to the width of the distribution between its 50th and 16th percentiles if the comparison is made towards a lower value, and between its 50th and 84th percentiles if the comparison is made towards a higher value. In a comparison, the σ values always refer to those belonging to the distribution that includes TDSL.

5.1 Cosmological inference from strong lensing alone

We first present our values for the cosmological parameters that can be inferred from TDSL alone. We use the time-delay distance likelihoods analytically expressed with a skewed lognormal distribution:

$$P(D_{\Delta t}) = \frac{1}{\sqrt{2\pi}(x-\lambda_D)\sigma_D}\exp\left[-\frac{(\ln(x-\lambda_D)-\mu_D)^2}{2\sigma_D^2}\right], \qquad (3)$$

---

[6] The time-delay distance measurement of RXJ1131−1231 from Suyu et al. (2014) that includes a composite model for the lens was not blind, whereas the first measurement of this same lens from Suyu et al. (2013) was done blindly.

[7] We use the Planck chains designated by 'plikHM_TT_lowTEB' that uses the baseline high-L Planck power spectra and low-L temperature and LFI polarization.

[8] http://pla.esac.esa.int/pla/#cosmology

Table 2. Description of the cosmological models considered in this work. *WMAP* refers to the constraints given in the *WMAP* Data Release 9. Planck refers either to the constraints from Planck 2015 Data Release alone, or combined with CMBL, BAO and/or JLA. See Section 5 for details.

Model name	Description
UH_0	Flat $-$ Λ CDM cosmology $\Omega_{\rm m} = 1 - \Omega_\Lambda = 0.32$ H_0 uniform in [0, 150]
UΛCDM	Flat $-$ Λ CDM cosmology $\Omega_{\rm m} = 1 - \Omega_\Lambda$ H_0 uniform in [0, 150] $\Omega_{\rm m}$ uniform in [0, 1]
UwCDM	Flat $-$ wCDM cosmology H_0 uniform in [0, 150] $\Omega_{\rm de}$uniform in [0, 1] w uniform in [$-$2.5, 0.5]
UoΛCDM	Non $-$ flat $-$ Λ CDM cosmology $\Omega_{\rm m} = 1 - \Omega_\Lambda - \Omega_{\rm k} > 0$ H_0 uniform in [0, 150] $\Omega_\Lambda$uniform in [0, 1] $\Omega_{\rm k}$ uniform in [$-$0.5, 0.5]
oΛCDM	Non $-$ flat $-$ Λ CDM cosmology *WMAP*/Planck for $\{H_0, \Omega_\Lambda, \Omega_{\rm m}\}$ $\Omega_{\rm k} = 1 - \Omega_\Lambda - \Omega_{\rm m}$
N$_{\rm eff}\Lambda$CDM	Flat $-$ Λ CDM cosmology *WMAP*/Planck for $\{H_0, \Omega_\Lambda, {\rm N_{eff}}\}$
m$_\nu\Lambda$CDM	Flat $-$ Λ CDM cosmology *WMAP*/Planck for $\{H_0, \Omega_\Lambda, \Sigma {\rm m}_\nu\}$
wCDM	Flat $-$ wCDM cosmology Planck for $\{H_0, w, \Omega_{\rm de}\}$
N$_{\rm eff}$m$_\nu\Lambda$CDM	Flat $-$ Λ CDM cosmology Planck for $\{H_0, \Omega_\Lambda, \Sigma {\rm m}_\nu, {\rm N_{eff}}\}$
owCDM	Open Λ CDM cosmology Planck for $\{H_0, \Omega_{\rm de}, \Omega_{\rm k}, w\}$

where $x = D_{\Delta t}/(1\,{\rm Mpc})$. We recall in Table 3 the lens and source redshifts of the three strong lenses as well as the parameters μ_D, σ_D and λ_D describing their respective time-delay distance distributions.

5.1.1 *Combination of three lenses*

Before carrying out a joint analysis of our three lens systems, we first perform a quantitative check that our three lenses can be combined without any loss of consistency. For that purpose, we compare their time-delay distance likelihood functions in the full cosmological parameter space, and measure the degree to which they overlap. Following Marshall, Rajguru & Slosar (2006) and Suyu et al. (2013), we compute the Bayes factor F, or evidence ratio, in favour of a simultaneous fit of the lenses using a common set of cosmological parameters. When comparing three data sets d_1, d_2 and d_3, we can either assume the hypothesis H$^{\rm global}$ that they can be represented using a common global set of cosmological parameters, or the hypothesis H$^{\rm ind}$ that at least one data set is better represented using another independent set of cosmological parameters. We stress that this latter model would make sense if there was a systematic error present that led to a vector offset in the inferred cosmological parameters. To parametrize this offset vector with no additional information would take as many nuisance parameters as there are dimensions in the cosmological parameter space; assigning uninformative uniform prior PDFs to each of the offset components is equivalent to using a complete set of independent cosmological parameters for the outlier 5 data set.

The Bayes factor, that makes the H$^{\rm global}$ hypothesis F times more probable than H$^{\rm ind}$ can be computed as follows:

$$F = \frac{P(d_1, d_2, d_3|{\rm H^{global}})}{P(d_1|{\rm H^{ind}}){\rm P}(d_2|{\rm H^{ind}}){\rm P}(d_3|{\rm H^{ind}})}. \quad (4)$$

A Bayes factor F significantly larger than 1 indicates that the considered data sets can be consistently combined. In the present case, considering three lenses with known time-delay distance likelihoods L_1, L_2 and L_3, the Bayes factor becomes

$$F_{1\cup 2\cup 3} = \frac{\langle L_1 L_2 L_3\rangle}{\langle L_1\rangle\langle L_2\rangle\langle L_3\rangle}, \quad (5)$$

where angle brackets denote averages over our ensembles of prior samples. We can also compare the likelihoods pair by pair (1-versus-1) as in equation 27 of Suyu et al. (2013) and then combine each pair with the third likelihood (2-versus-1):

$$F_{12\cup 3} = \frac{\langle L_1 L_2 L_3\rangle}{\langle L_1 L_2\rangle\langle L_3\rangle}. \quad (6)$$

This last equation allows us to check that the lenses can also be well combined pair-wise, and that none of them is inconsistent with the two others considered together. We compute the Bayes Factors $F_{1\cup 2\cup 3}$ and all the possible 1-versus-1 and 2-versus-1 permutations in the uniform cosmologies UH_0, UΛCDM, UwCDM and UoΛCDM. We find that all the combinations are in good agreement, the only exception being for the pair B1608+656$\cup$RXJ1131$-$1231, which is only marginally consistent in the UoΛCDM cosmology ($F_{1\cup 2}$=1.1). Considering the likelihoods individually, the three lenses are in excellent agreement, with a Bayes Factor $F_{1\cup 2\cup 3} =$ 21.3 in UH_0, $F_{1\cup 2\cup 3} = 14.2$ in UΛCDM, $F_{1\cup 2\cup 3} = 18.9$ in UwCDM and $F_{1\cup 2\cup 3} = 10.8$ in UoΛCDM. We conclude that the time-delay likelihoods of our three lenses can be combined without any loss of consistency.

5.1.2 *Constraints in uniform cosmologies*

Fig. 4 presents the marginalized posterior PDF for H_0 in the cosmological models using uniform priors. Our baseline model, UH_0, has a uniform prior on H_0 in the range [0, 150] km s^{-1} Mpc^{-1}, a matter density fixed at $\Omega_{\rm m} = 0.32$ from the most recent Planck results (Planck Collaboration XIII 2016a), zero curvature $\Omega_{\rm k} = 0$ and consequently a fixed cosmological constant. This model has only one free parameter. From left to right in the figure, we present this UH_0 cosmology, and then three models that have two or three

Table 3. Parameters of the three strong lenses used in our analysis. μ_D, σ_D and λ_D are related to the analytical fit of the time-delay distance probability function (see equation 3).

Name	Reference	$z_{\rm d}$	$z_{\rm s}$	μ_D	σ_D	λ_D
B1608+656	Suyu et al. (2010)	0.6304	1.394	7.0531	0.228 24	4000.0
RXJ1131$-$1231	Suyu et al. (2014)	0.295	0.654	6.4682	0.205 60	1388.8
HE 0435$-$1223	H0LiCOW Paper IV	0.4546	1.693	7.5793	0.103 12	653.9

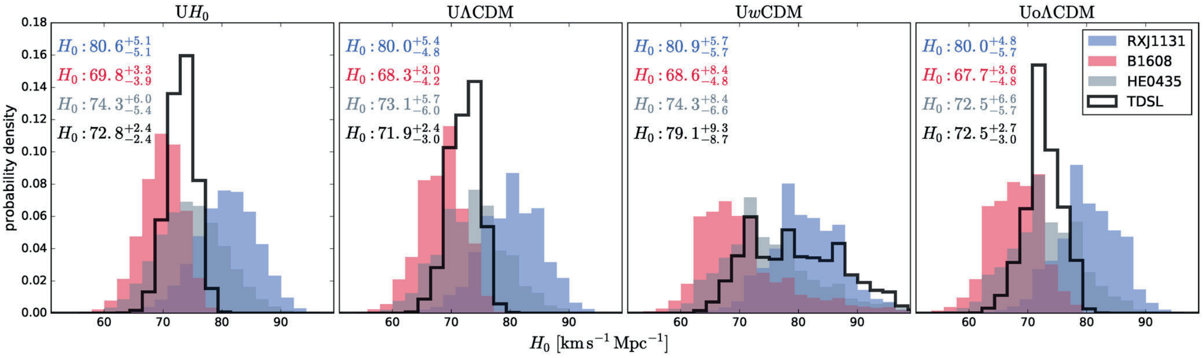


Figure 4. Marginalized posterior probability distributions for H_0 in the UH_0, UΛCDM, UwCDM and UoΛCDM cosmologies using the constraints from the three strong lenses B1608+656, RXJ1131−1231 and HE 0435−1223. The overlaid histograms present the distributions for each individual strong lens (ignoring the other two data sets), and the solid black line corresponds to the distribution resulting from the joint inference from all three data sets (TDSL). The quoted values of H_0 in the top-left corner of each panel are the median, 16th and 84th percentiles.

free parameters (H_0 plus one or two others): the UΛCDM cosmology where we allow Ω_m to vary with uniform prior; the UwCDM cosmology with a free Ω_de and a free time-independent dark energy equation of state w, both with uniform priors; and finally the UoΛCDM cosmology, that relaxes the constraint on the curvature Ω_k and allows both this and Ω_Λ to vary with uniform priors. Table 2 summarizes the constraints and priors for these four models. We quote in each panel the corresponding median and 1σ uncertainties of H_0. In the UH_0 cosmology, combining the three lenses yields a value $H_0 = 72.8 \pm 2.4$ km s^{-1} Mpc^{-1}, with 3.3 per cent precision. When relaxing the constraint on Ω_m in UΛCDM (and thus being completely independent of any other measurement), we obtain $H_0 = 71.9^{+2.4}_{-3.0}$ km s^{-1} Mpc^{-1}, with 3.8 per cent precision. These two estimates are, respectively, 2.5σ and 1.7σ higher than the most recent Planck measurement in a flat-ΛCDM universe ($H_0 = 66.93 \pm 0.62$ km s^{-1} Mpc^{-1}; Planck Collaboration XIII 2016a), in excellent agreement with the most recent results using distance ladders ($H_0 = 73.24 \pm 1.74$ km s^{-1} Mpc^{-1}; Riess et al. 2016), and compatible with other local estimates (see e.g. Bonamente et al. 2006; Freedman et al. 2012; Sorce, Tully & Courtois 2012; Gao et al. 2016). Whether the tension between the local and cosmological measurements of H_0 comes from systematic errors or hints at new physics beyond flat ΛCDM is currently a hot topic of discussion in the community (see e.g. Efstathiou 2014; Rigault et al. 2015; Spergel, Flauger & Hložek 2015; Addison et al. 2016; Di Valentino, Melchiorri & Silk 2016; Planck Collaboration XIII 2016a; Riess et al. 2016, and references therein).

Intriguingly, we note that the H_0 values yielded by each system individually get larger for lower lens redshifts. So far, we cannot state if this comes from a simple statistical fluke, an unknown systematic error or hints towards an unaccounted physical property. The addition of two more lenses from the H0LiCOW sample will certainly help us in that regard.

Fig. 5 presents the two-dimensional 95 per cent credible regions of the cosmological parameters in the UwCDM, UoΛCDM and UΛCDM cosmologies for each lens individually and for their combination (TDSL). TDSL is primarily sensitive to H_0, and the tilt in the $H_0 - \Omega_\Lambda$, $H_0 - w$ and $H_0 - \Omega_\mathrm{k}$ planes illustrates its weak sensitivity to the dark energy density, dark energy equation of state and curvature density, respectively. TDSL alone agrees both with a flat universe and a cosmological constant, although on the latter the credible region extends deeply into the phantom dark energy domain ($w < -1$). In the UwCDM cosmology, the correlation between H_0 and w is more prominent than in the other models, leading to a larger dispersion of the H_0 distribution in the corresponding panel of Fig. 4. This dispersion is more prominent for values of $w < -1$, since in such cases the variation of the density of dark

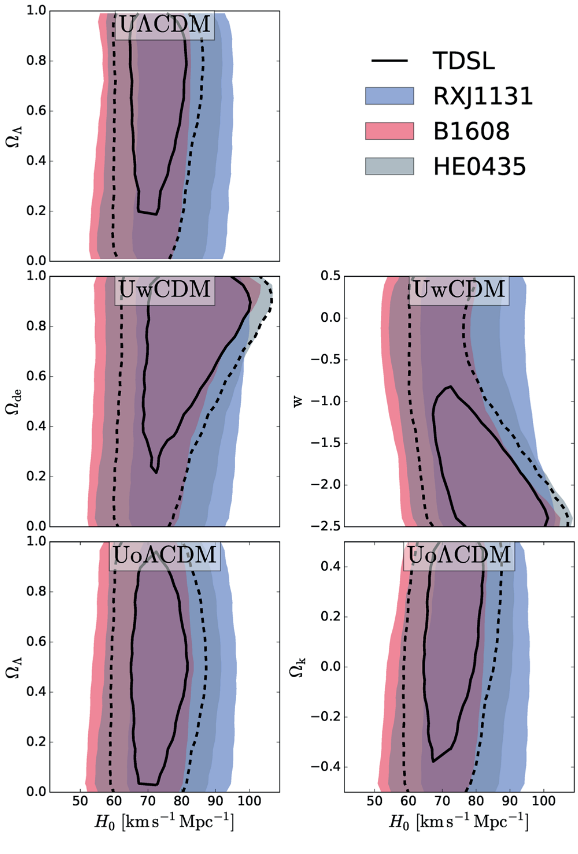


Figure 5. Comparison of the three strong lenses in the UΛCDM (top), UwCDM (middle) and UoΛCDM (bottom) cosmologies. The coloured overlays delimit the 95 per cent credible region for B1608+656, RXJ1131−1231 and HE 0435−1223. The solid and dashed black lines draw the contours of the 68.3 and 95 per cent credible regions, respectively, for the combination of the three lenses.

Table 4. Summary of the cosmological parameters constraints for the models detailed in Table 2. H_0 units are km s^{-1} Mpc^{-1}, Σm_ν units are eV. The quoted values are the median, 16th and 84th percentiles, except for Σm_ν where we quote the 95 per cent upper bound of the probability distribution. The empty slots occur when no prior samples are provided by the Planck team.

	UH_0	UΛCDM		UwCDM			UoΛCDM		
	H_0	H_0	Ω_Λ	H_0	Ω_{de}	w	H_0	Ω_Λ	Ω_k
TDSL	$72.8^{+2.4}_{-2.4}$	$71.9^{+2.4}_{-3.0}$	$0.62^{+0.24}_{-0.35}$	$79.1^{+9.3}_{-8.7}$	$0.72^{+0.19}_{-0.34}$	$-1.79^{+0.94}_{-0.49}$	$72.5^{+2.7}_{-3.0}$	$0.51^{+0.28}_{-0.30}$	$0.1^{+0.3}_{-0.3}$

	oΛCDM				wCDM		
	H_0	Ω_m	Ω_Λ	Ω_k	H_0	Ω_{de}	w
TDSL+*WMAP*	$73.0^{+2.3}_{-2.5}$	$0.25^{+0.02}_{-0.02}$	$0.74^{+0.02}_{-0.02}$	$0.005^{+0.005}_{-0.005}$	$76.5^{+4.6}_{-3.9}$	$0.76^{+0.02}_{-0.02}$	$-1.24^{+0.16}_{-0.20}$
TDSL+Planck (1)	$69.2^{+1.4}_{-2.2}$	$0.30^{+0.02}_{-0.02}$	$0.70^{+0.01}_{-0.01}$	$0.003^{+0.004}_{-0.006}$	$79.0^{+4.4}_{-4.2}$	$0.77^{+0.02}_{-0.03}$	$-1.38^{+0.14}_{-0.16}$
(1)+BAO	$68.0^{+0.7}_{-0.7}$	$0.31^{+0.01}_{-0.01}$	$0.69^{+0.01}_{-0.01}$	$0.001^{+0.003}_{-0.003}$	$69.6^{+1.8}_{-1.7}$	$0.70^{+0.01}_{-0.01}$	$-1.08^{+0.07}_{-0.08}$
(1)+BAO+JLA	$68.1^{+0.7}_{-0.7}$	$0.31^{+0.01}_{-0.01}$	$0.69^{+0.01}_{-0.01}$	$0.001^{+0.003}_{-0.003}$	$68.8^{+1.0}_{-1.0}$	$0.70^{+0.01}_{-0.01}$	$-1.04^{+0.05}_{-0.05}$

	N_{eff}ΛCDM			m_νΛCDM		
	H_0	Ω_Λ	N_{eff}	H_0	Ω_Λ	Σm_ν (eV)
TDSL+*WMAP*	$73.2^{+2.2}_{-2.4}$	$0.72^{+0.02}_{-0.03}$	$3.86^{+0.73}_{-0.71}$	$70.7^{+1.9}_{-1.9}$	$0.73^{+0.02}_{-0.02}$	≤ 0.393
TDSL+Planck (1)	$71.0^{+2.0}_{-2.0}$	$0.71^{+0.01}_{-0.01}$	$3.45^{+0.23}_{-0.24}$	$68.1^{+1.1}_{-1.2}$	$0.70^{+0.01}_{-0.02}$	≤ 0.199
(1)+BAO	$69.6^{+1.4}_{-1.3}$	$0.70^{+0.01}_{-0.01}$	$3.34^{+0.21}_{-0.21}$	$67.9^{+0.6}_{-0.6}$	$0.69^{+0.01}_{-0.01}$	≤ 0.182
(1)+BAO+CMBL				$67.9^{+0.6}_{-0.7}$	$0.69^{+0.01}_{-0.01}$	≤ 0.216

	$N_{eff}m_\nu$ΛCDM				owCDM			
	H_0	Ω_Λ	N_{eff}	Σm_ν (eV)	H_0	Ω_{de}	Ω_k	w
TDSL+Planck (1)	$70.8^{+2.0}_{-2.1}$	$0.71^{+0.02}_{-0.02}$	$3.44^{+0.24}_{-0.24}$	≤ 0.274	$88.4^{+5.9}_{-7.2}$	$0.83^{+0.02}_{-0.03}$	$-0.010^{+0.003}_{-0.003}$	$-2.10^{+0.34}_{-0.41}$
(1)+CMBL	$70.8^{+2.1}_{-2.1}$	$0.71^{+0.02}_{-0.02}$	$3.44^{+0.25}_{-0.24}$	≤ 0.347	$77.9^{+5.0}_{-4.2}$	$0.77^{+0.03}_{-0.03}$	$-0.003^{+0.004}_{-0.004}$	$-1.37^{+0.18}_{-0.23}$
(1)+BAO+CMBL					$70.0^{+2.1}_{-1.7}$	$0.71^{+0.02}_{-0.02}$	$-0.000^{+0.004}_{-0.003}$	$-1.07^{+0.09}_{-0.10}$

energy becomes larger at low redshifts. Since our measurements are performed at the redshift of the lenses we observe, going back to redshift zero and H_0 produces the degeneracy with w.

This highlights the fact that our cosmological inferences in this cosmology are more sensitive to the prior range we choose. Thus, the resulting parameter values must be considered as indicative of a trend rather than as absolute measurements. We summarize our values for H_0, Ω_k, w and Ω_m from TDSL alone in the top section of Table 4.

5.2 Constraining cosmological models beyond ΛCDM

We now investigate how strong lensing can help constrain cosmological models beyond standard flat ΛCDM, when combined with other cosmological probes. We demonstrated in Section 5.1 and Fig. 5 that TDSL is only weakly dependent on the matter density, the dark energy density, the dark energy equation of state and the curvature. However, the cosmological parameter degeneracies for TDSL are such that the combination of TDSL with other probes can rule out large areas of parameter space. Following the motivations presented in Planck Collaboration XIII (2016a) for extensions to the base ΛCDM model, we present in the following a selection of models where we combine TDSL with the results from *WMAP*, Planck, Planck+BAO, Planck+BAO+CMBL and Planck+BAO+JLA when available. Figs 6 and 7 present the results. Note that we have smoothed the contours of the credible regions after importance sampling with a Gaussian filter due to the sparsity of the *WMAP* and Planck MCMC chains, checking that the 95 per cent credible regions do indeed contain approximately 95 per cent of the importance weight.

5.2.1 *One-parameter extensions*

We first consider one-parameter extensions to the standard model, where we relax the constraints on one additional cosmological parameter from flat ΛCDM. We present in Fig. 6 the two-dimensional marginalized parameter space for a selection of cosmological models for which the impact of TDSL is the most meaningful.

In the oΛCDM model, we consider a non-flat universe with $\Omega_k \neq 0$. In the N_{eff}ΛCDM model, we consider a variable effective number of relativistic neutrino species N_{eff} with a fixed total mass of neutrinos $\Sigma m_\nu = 0.06$ eV. In the m_νΛCDM model, we consider a variable Σm_ν with a fixed $N_{eff} = 3.046$. Finally, in the wCDM model we consider a time-invariant dark energy equation of state w. A detailed description of these models is given in Table 2.

For each probe, or combination of probes, we draw the 95 per cent credible region contours as coloured lines. When combined with TDSL, the updated credible region is displayed as a filled area.

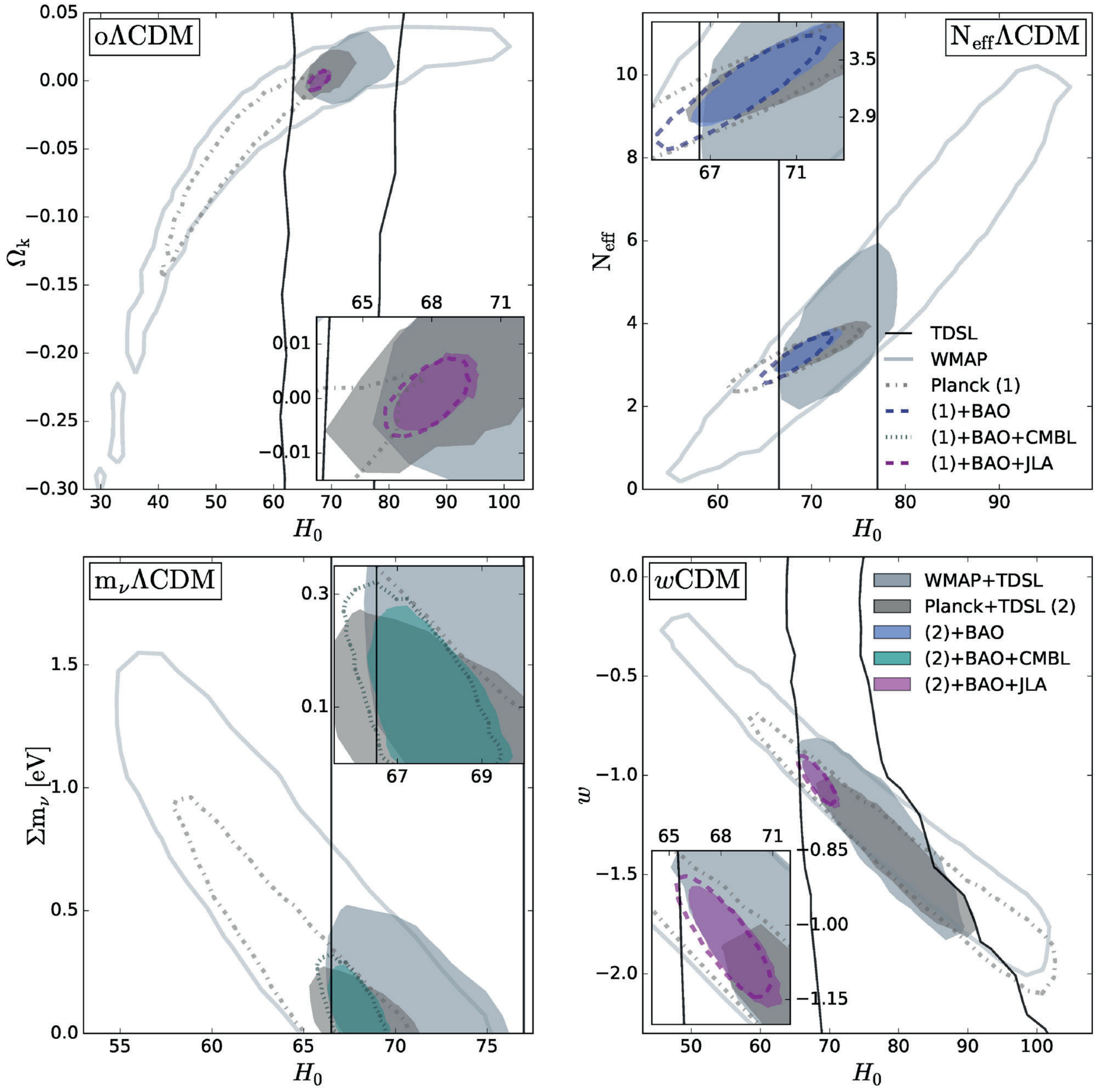


Figure 6. Cosmological constraints in one-parameter extensions to ΛCDM. We consider a non-flat universe with variable curvature Ω_k (top-left), a variable effective number of relativistic neutrino species N_{eff} (top-right), a variable total mass of neutrino species Σm_ν (bottom-left, in eV) and a variable time-invariant dark energy equation of state w (bottom-right). The filled regions and coloured lines delimit the marginalized 95 per cent credible regions (consistently smoothed due to the sparsity of the samples from the available MCMC chains) with and without the constraints from TDSL, respectively. The different colours represent the constraints from *WMAP*, Planck, Planck+CMBL, Planck+BAO, Planck+CMBL+BAO and Planck+BAO+JLA. The solid black lines delimit the 95 per cent credible region for TDSL alone in the corresponding uniform cosmology with no additional information.

When importance sampling using priors based on other probes, it is important to verify that the respective constraints in the parameter space overlap. If they do not, the probes considered may not be efficiently combined. With this in mind, we plot in each cosmology the 95 per cent credible region for TDSL only (and uniform priors) as thin solid black lines. We note that in all one-parameter extensions presented here, the 2D marginalized TDSL and Planck 95 per cent credible regions at least partially overlap, although in the oΛCDM and m_νΛCDM cosmologies, the 1D marginalized posterior mean value for H_0 from TDSL lies outside the 95 per cent credible region of Planck. We consider the overlaps to be sufficient to justify our importance sampling TDSL with Planck, but emphasize that the joint constraints must be interpreted cautiously. *WMAP* and Planck constraints are in agreement with each other, this being at least partly due to the large parameter space covered in the credible region of *WMAP*. This also results in a much wider overlap with the TDSL 95 per cent credible regions.

We summarize our inferred values for H_0 and other cosmological parameters of each cosmology in Table 4. In the oΛCDM cosmology, both *WMAP*+TDSL and Planck+TDSL are consistent with a

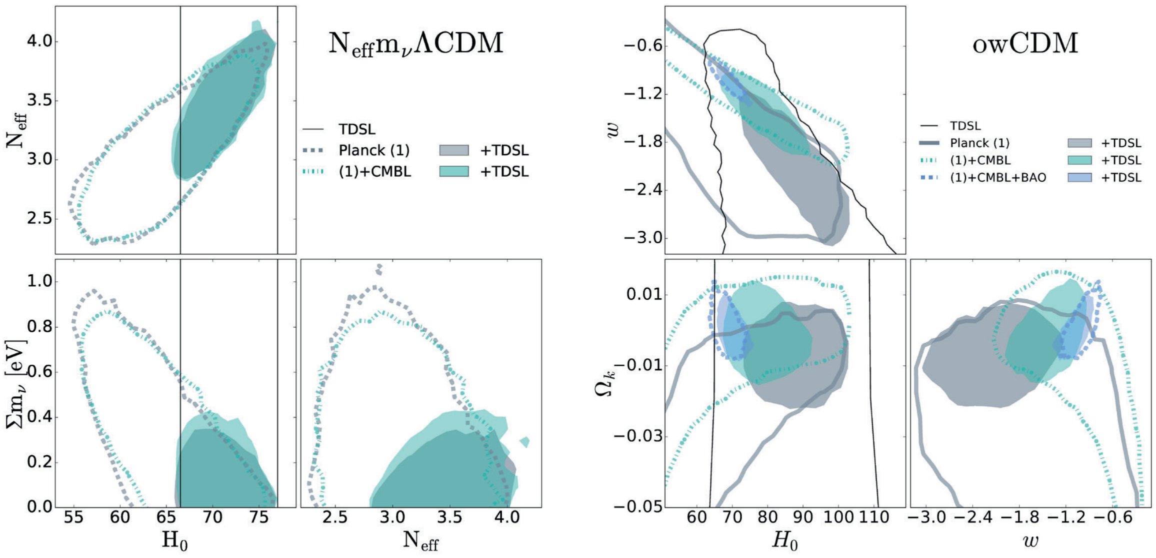


Figure 7. Cosmological constraints in two-parameter extensions to ΛCDM. We consider a flat universe with variable effective number of relativistic neutrino species $N_{\rm eff}$ and total mass of neutrinos Σm_ν (left), and an open universe with variable dark energy equation of state parameters w (right). The coloured lines and filled areas are the same as in Fig. 6, and show marginalized 95 per cent credible regions. The TDSL contours in the owCDM cosmology are computed using uniform priors on $\Omega_{\rm k}$ [−0.5, 0.5] and w [−2.5, 0.5].

flat universe. The constraints on $\Omega_{\rm k}$ from Planck+TDSL are approximately twice as large as those from Planck+BAO+JLA. In the $m_\nu\Lambda$CDM cosmology, the upper bound of the sum of the neutrino masses Σm_ν from *WMAP*+TDSL is approximately twice as large as the prediction from Planck+TDSL. The addition of TDSL lowers the upper bound from Planck+BAO by about 5 per cent. The joint constraint from Planck+BAO+TDSL yields $\Sigma m_\nu \leq 0.182$ eV with 95 per cent probability. In the $N_{\rm eff}\Lambda$CDM cosmology, both *WMAP*+TDSL and Planck+TDSL suggest an effective number of relativistic neutrino species higher than the standard cosmological prediction of $N_{\rm eff} = 3.046$. The Planck+TDSL value is similar in precision to Planck+BAO, yet the two values are in mild tension, the former being 1.3σ higher. The combination of Planck+BAO+TDSL yields $N_{\rm eff} = 3.34 \pm 0.21$, also in mild tension with the standard cosmological prediction. In the wCDM cosmology, Planck+TDSL points towards $w = -1.38^{+0.14}_{-0.16}$, a result in tension with a cosmological constant ($w = -1$) at a 2.3σ level. This value is lower than other values for phantom dark energy reported in the literature (see e.g. Freedman et al. 2012; Collett & Auger 2014). With *WMAP*+TDSL we find $w = -1.24^{+0.16}_{-0.20}$, consistent with the previous measurement from our group using just B1608+656 and RXJ1131−1231 combined with *WMAP* of $w = -1.14^{+0.17}_{-0.20}$ (Suyu et al. 2013).

5.2.2 *Two-parameter extensions*

We now consider cosmological models where we relax the priors on two cosmological parameters from flat ΛCDM. Following the discussion of Section 5.2.1 where we noted that the individual TDSL and Planck 95 per cent credible regions only partially overlap, we consider here two cosmological models that reduce the tension between these two probes. First, we consider the $N_{\rm eff}m_\nu\Lambda$CDM model, where both the effective number of relativistic neutrino species $N_{\rm eff}$ as well as their total mass Σm_ν are allowed to vary. Secondly, we consider the owCDM model where we relax the constraints on both the curvature, $\Omega_{\rm k}$, and the dark energy equation of state parameter w simultaneously. For the owCDM model, the Planck team does not publicly provide MCMC chains. We therefore generate additional chains using the publicly available Planck cosmological likelihood function, *plik* (Planck Collaboration XVIII 2016c). Temperature power spectra were computed using the Cosmic Linear Anisotropy Solving System Boltzman code (CLASS; Blas, Lesgourgues & Tram 2011; Lesgourgues & Tram 2014) and MCMC chains were generated with the MontePython sampler (Audren et al. 2013).

Fig. 7 presents the two-dimensional marginalized 95 per cent credible regions for these two models, and the bottom of Table 4 reports the 1D marginalized posterior median values and 1σ uncertainties of the corresponding model parameters. We note that this time, the TDSL and Planck 95 per cent credible regions are in much better agreement than in the one-parameter extension models. In the $N_{\rm eff}m_\nu\Lambda$CDM cosmology, Planck alone and Planck+CMBL are in agreement with the standard cosmological prediction of $N_{\rm eff}$, yet the constraints are rather large. When adding TDSL, the constraints are strongly tightened and we obtain $N_{\rm eff} = 3.44 \pm 0.24$, in mild tension with the standard cosmological prediction of $N_{\rm eff} = 3.046$. Similarly, the constraints on the maximum neutrino mass are tightened by a factor $\simeq$3 when adding TDSL, yielding $\Sigma m_\nu \leq 0.274$ eV with 95 per cent probability. In the owCDM cosmology, Planck+CMBL+TDSL yields $\Omega_{\rm k} = -0.003^{+0.004}_{-0.004}$, in good agreement with Planck+CMBL+BAO and in favour of a flat universe. However, a tension in the dark energy equation of state w still remains, as Planck+CMBL+TDSL yields $w = -1.37^{-0.23}_{+0.18}$, 2$\sigma$ lower than the cosmological constant prediction.

6 CONCLUSIONS

Using multiple telescopes in the Southern and Northern hemispheres, we have monitored the quadruple-imaged strong gravitational lens HE 0435−1223 for 13 yr with an average cadence of one observing epoch every 3.6 d. We analyse the imaging data using the MCS deconvolution algorithm (Magain et al. 1998) on a total of 876 observing epochs to produce the light curves of the four lensed images, with an rms photometric precision of 10 mmag on the brightest quasar image.

We measured the time delays between each pair of lensed images using the free-knot spline technique and the regression difference technique from the PYCS package (Tewes et al. 2013a). Our uncertainty estimation involves the generation of synthetic light curves that closely mimic the intrinsic and extrinsic features of the real data. To test the robustness of our measurements, we vary parameters such as the number of knots in the splines, the initial parameters used for the deconvolution photometry and the length of the considered light curves. The two curve shifting techniques agree well with each other both on the point estimation of the delays and on the estimated uncertainty. The smallest relative uncertainty, of 6.5 per cent, is obtained for the A-D pair of images. For this pair involving image A, our present measurement is twice as precise as the earlier result by Courbin et al. (2011).

In H0LiCOW Paper IV, we use our new COSMOGRAIL time delays for HE 0435−1223 to compute its time-delay distance. Very importantly, this is done in a blind way with respect to the inference of cosmological parameters. In this paper, we combine the time-delay distance likelihoods from HE 0435−1223 with the published ones from B1608+656 and RXJ1131−1231 to create a TDSL probe. We also combine the latter with other cosmological probes such as *WMAP*, Planck, BAO and JLA to constrain cosmological parameters for a large range of extended cosmological models. Our main conclusions are as follows.

(i) TDSL alone is weakly sensitive to the matter density, $\Omega_{\rm m}$, curvature, $\Omega_{\rm k}$ and dark energy density $\Omega_{\rm de}$ and equation of state w. Its primary sensitivity to H_0 allows us to break degeneracies of CMB probes in extended cosmological models.

(ii) In a flat-ΛCDM cosmology with uniform priors on H_0 and $\Omega_{\rm m}$, TDSL alone yields $H_0 = 71.9^{+2.4}_{-3.0}\,{\rm km\,s^{-1}\,Mpc^{-1}}$. When enforcing $\Omega_{\rm m}$=0.32 from the most recent Planck results, we find $H_0 = 72.8 \pm 2.4\,{\rm km\,s^{-1}\,Mpc^{-1}}$. These results are in excellent agreement with the most recent measurements using the distance ladder, but are in tension with the CMB measurements from Planck.

(iii) In a non-flat-ΛCDM cosmology, we find, using TDSL and Planck, H_0 =$69.2^{+1.4}_{-2.2}\,{\rm km\,s^{-1}\,Mpc^{-1}}$ and $\Omega_{\rm k} = 0.003^{+0.004}_{-0.006}$ in agreement with a flat universe.

(iv) In a flat-wCDM cosmology in combination with Planck, we find a 2.3σ tension from a cosmological constant in favour of a phantom form of dark energy. Our joint constraints in this cosmology are $H_0 = 79.0^{+4.4}_{-4.2}\,{\rm km\,s^{-1}\,Mpc^{-1}}$, $\Omega_{\rm de} = 0.77^{+0.02}_{-0.03}$ and $w = -1.38^{+0.14}_{-0.16}$.

(v) In a flat-m$_\nu\Lambda$CDM cosmology, in combination with Planck and BAO we tighten the constraints on the maximum mass of neutrinos to $\Sigma{\rm m}_\nu \leq$0.182 eV, while removing the tension in H_0.

(vi) In a flat-N$_{\rm eff}\Lambda$CDM cosmology, in combination with Planck and BAO we find N$_{\rm eff}$ =3.34 ± 0.21, i.e. 1.3σ higher than the standard cosmological value. This mild tension remains when the constraints on both $\Sigma{\rm m}_\nu$ and N$_{\rm eff}$ are relaxed.

(vii) In a owCDM cosmology, in combination with Planck and CMBL, we find $H_0 = 77.9^{+5.0}_{-4.2}\,{\rm km\,s^{-1}\,Mpc^{-1}}$, $\Omega_{\rm de} = 0.77^{+0.03}_{-0.03}$, $\Omega_{\rm k} = -0.003^{+0.004}_{-0.004}$ and $w = -1.37^{+0.18}_{-0.23}$. Similarly to the o$\Lambda$CDM and wCDM cosmologies, we are in good agreement with a flat universe and in tension with a cosmological constant, respectively.

We emphasize that despite reporting parameter constraints for a large variety of cosmological models beyond ΛCDM, we choose not to comment on whether a particular model is favoured over the others. Such an exercise would require a well-motivated choice of priors for these models, which is not within the scope of this work.

The combined strengths of our H0LiCOW lens modelling and COSMOGRAIL monitoring indicate that quasar time-delay cosmography is now a mature field, producing precise and accurate inferences of cosmological parameters, that are independent of any other cosmological probe. Still, our results can be improved in at least four ways.

(i) Continuing to enlarge the sample. Two more objects with excellent time-delay measurements as well as high-resolution imaging and spectroscopic data remain to be analysed in the H0LiCOW project (see H0LiCOW Paper I). When completed, H0LiCOW is expected to provide a measurement of H_0 to better than 3.5 per cent in a non-flat-ΛCDM universe with flat priors on $\Omega_{\rm m}$ and Ω_Λ. Data of quality comparable to those obtained for H0LiCOW are in the process of being obtained for another four systems with measured time delays from COSMOGRAIL (HST-GO-14254; PI: Treu). Meanwhile, current and planned wide field imaging surveys such as DES, KiDS, HSC, LSST, *Euclid* and WFIRST, should discover hundreds of new gravitational lens systems suitable for time-delay cosmography (Oguri & Marshall 2010). For example, the dedicated search in the Dark Energy Survey STRIDES[9] has already confirmed two new lenses from the Year1 data (Agnello et al. 2016).

(ii) Improve the lens modelling accuracy. The tests carried out in our current (H0LiCOW Paper IV) and past work (Suyu et al. 2014), the good internal agreement between the three measured systems (Section 5.1), and independent analysis based on completely independent codes (Birrer et al. 2016), show that our lens models are sufficiently complex given the currently available data. However, as the number of systems being analysed grows, random uncertainties in the cosmological parameters will fall, and residual systematic uncertainties related to degeneracies inherent to gravitational lensing will need to be investigated in more detail. Following the work of Xu et al. (2016), detailed hydro *N*-body simulations of lensing galaxies in combination with ray-shooting can be used to evaluate the impact of the lensing degeneracies on cosmological results in view of future observations with the JWST or 30-m class ground-based telescopes with adaptive optics, and to drive development of improved lens modelling techniques and assumptions appropriate to the density structures we expect.

(iii) Improve the absolute mass calibration. Spatially resolved 2D kinematics of the lens galaxies, to be obtained either with JWST and with integral field spectrographs mounted on large ground-based telescopes with adaptive optics, should further improve both the precision for each system and our ability to test for residual systematics, including those arising from the mass sheet and SPT (Schneider & Sluse 2013, 2014; Unruh et al. 2016). The same data should also allow us to use constraints from the stellar mass or mass profile of lens galaxies as attempted in Courbin et al. (2011) with slit spectroscopy. Alternatively, the MSD can be lifted if the absolute luminosity of the source is known (Falco et al. 1985), which is the case for lensed standard candles (see e.g. Goobar et al. 2016, that report the first discovery of a lensed Type Ia Supernovae). However, such configurations happens far less often than lensed quasars.

(iv) Measuring time delays with the current photometric precision and time sampling of monitoring data requires long and time-consuming campaigns, and is currently not possible for hundreds of objects. Increasing the monitoring efficiency is possible, by

---

[9] strides.astro.ucla.edu

catching extremely small (mmag) and fast (days) variations in the quasar light curves. Such data can be obtained with daily observations with 2-m class telescopes in good seeing conditions, a project that will be implemented in the context of the extended COSMOGRAIL programme (eCOSMOGRAIL; Courbin et al. 2016, in preparation) to measure quasar time delays in only one or two observing seasons. Furthermore, in the long run, LSST should be able to provide sufficiently accurate time delays for hundreds of systems from the survey data itself (Liao et al. 2015), and enable subpercent precision on H_0 in the next decade (Treu & Marshall 2016).

ACKNOWLEDGEMENTS

We thank Adriano Agnello, Roger Blandford, Geoff ChihFan Chen, Xuheng Ding, Yashar Hezaveh, Kai Liao, John McKean, Danka Paraficz, Olga Tihhonova and Simona Vegetti for their contributions to the H0LiCOW project. We thank the anonymous referee for his or her comments. We also thank all the observers at the Euler, SMARTS, Mercator and Maidanak telescopes who participated in the queue-mode observations. H0LiCOW and COSMOGRAIL are made possible, thanks to the continuous work of all observers and technical staff obtaining the monitoring observations, in particular at the Swiss Euler telescope at La Silla Observatory. The Euler telescope is supported by the Swiss National Science Foundation. We are grateful to Thomas Tram for expert help in implementing CLASS and MontePython to create MCMC chains for non-standard cosmologies. Numerical computations for the CMB likelihoods were done on the Sciama High Performance Compute (HPC) cluster which is supported by the ICG, SEPNet and the University of Portsmouth. VB, FC and GM acknowledge the support of the Swiss National Science Foundation (SNSF). SHS acknowledges support from the Max Planck Society through the Max Planck Research Group. This work is supported in part by the Ministry of Science and Technology in Taiwan via grant MOST-103-2112-M-001-003-MY3. KCW is supported by an EACOA Fellowship awarded by the East Asia Core Observatories Association, which consists of the Academia Sinica Institute of Astronomy and Astrophysics, the National Astronomical Observatory of Japan, the National Astronomical Observatories of the Chinese Academy of Sciences and the Korea Astronomy and Space Science Institute. PJM acknowledges support from the US Department of Energy under contract number DE-AC02-76SF00515. CER and CDF were funded through the NSF grant AST-1312329, 'Collaborative Research: Accurate cosmology with strong gravitational lens time delays', and the *HST* grant GO-12889. DS acknowledges funding support from a *Back to Belgium* grant from the Belgian Federal Science Policy (BELSPO). MT acknowledges support by a fellowship of the Alexander von Humboldt Foundation and the DFG grant Hi 1495/2-1. SH acknowledges support by the DFG cluster of excellence 'Origin and Structure of the Universe' (www.universe-cluster.de). TT thanks the Packard Foundation for generous support through a Packard Research Fellowship, the NSF for funding through NSF grant AST-1450141, 'Collaborative Research: Accurate cosmology with strong gravitational lens time delays'. LVEK is supported in part through an NWO-VICI career grant (project number 639.043.308).

Based on observations made with the NASA/ESA *HST*, obtained at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. These observations are associated with programme #12889, #10158, #9744 and #7422. Support for programme #12889 was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555.

REFERENCES

Addison G. E., Huang Y., Watts D. J., Bennett C. L., Halpern M., Hinshaw G., Weiland J. L., 2016, ApJ, 818, 132
Agnello A., Sonnenfeld A., Suyu S. H., Treu T., Fassnacht C. D., Mason C., Bradač M., Auger M. W., 2016, MNRAS, 458, 3830
Anderson L. et al., 2012, MNRAS, 427, 3435
Audren B., Lesgourgues J., Benabed K., Prunet S., 2013, J. Cosmol. Astropart. Phys., 2, 001
Bennett C. L. et al., 2013, ApJS, 208, 20
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Betoule M. et al., 2013, A&A, 552, A124
Beutler F. et al., 2011, MNRAS, 416, 3017
Birrer S., Amara A., Refregier A., 2016, J. Cosmol. Astropart. Phys., 08, 020
Blackburne J. A., Kochanek C. S., Chen B., Dai X., Chartas G., 2014, ApJ, 789, 125
Blake C. et al., 2011, MNRAS, 418, 1707
Blas D., Lesgourgues J., Tram T., 2011, Astrophysics Source Code Library, record ascl:1106.020
Bonamente M., Joy M. K., LaRoque S. J., Carlstrom J. E., Reese E. D., Dawson K. S., 2006, ApJ, 647, 25
Bonvin V., Tewes M., Courbin F., Kuntzer T., Sluse D., Meylan G., 2016, A&A, 585, A88
Braibant L., Hutsemékers D., Sluse D., Anguita T., García-Vergara C. J., 2014, A&A, 565, L11
Cantale N., Courbin F., Tewes M., Jablonka. P., Meylan G., 2016, A&A, 589, A81
Chen G. C. F. et al., 2016, MNRAS, 462, 3457
Collett T. E., Auger M. W., 2014, MNRAS, 443, 969
Collett T. E. et al., 2013, MNRAS, 432, 679
Courbin F., Eigenbrod A., Vuissoz C., Meylan G., Magain P., 2005, in Mellier Y., Meylan G., eds, Proc. IAU Symp. 225, Gravitational Lensing Impact on Cosmology. Cambridge Univ. Press, Cambridge, p. 297
Courbin F. et al., 2011, A&A, 536, A53
Di Valentino E., Melchiorri A., Silk J., 2016, Phys. Lett. B, 761, 242
Dobler G., Fassnacht C. D., Treu T., Marshall P., Liao K., Hojjati A., Linder E., Rumbaugh N., 2015, ApJ, 799, 168
Efstathiou G., 2014, MNRAS, 440, 1138
Eigenbrod A., Courbin F., Dye S., Meylan G., Sluse D., Vuissoz C., Magain P., 2006a, A&A, 451, 747
Eigenbrod A., Courbin F., Meylan G., Vuissoz C., Magain P., 2006b, A&A, 451, 759
Eulaers E. et al., 2013, A&A, 553, A121
Falco E. E., Gorenstein M. V., Shapiro I. I., 1985, ApJ, 289, L1
Fassnacht C. D., Xanthopoulos E., Koopmans L. V. E., Rusin D., 2002, ApJ, 581, 823
Fassnacht C. D., Gal R. R., Lubin L. M., McKean J. P., Squires G. K., Readhead A. C. S., 2006, ApJ, 642, 30
Fassnacht C. D., Koopmans L. V. E., Wong K. C., 2011, MNRAS, 410, 2167
Freedman W. L., Madore B. F., Scowcroft V., Burns C., Monson A., Persson S. E., Seibert M., Rigby J., 2012, ApJ, 758, 24
Gao F. et al., 2016, ApJ, 817, 128
Giusarma E., Gerbino M., Mena O., Vagnozzi S., Ho S., Freese K., 2016, Phys. Rev. D, 94, 083522
Goobar A. et al., 2016, preprint (arXiv:1611.00014)
Greene Z. S. et al., 2013, ApJ, 768, 39
Heavens A., Jimenez R., Verde L., 2014, Phys. Rev. Lett., 113, 241302
Heymans C. et al., 2012, MNRAS, 427, 146
Hilbert S., White S. D. M., Hartlap J., Schneider P., 2007, MNRAS, 382, 121
Hilbert S., Hartlap J., White S. D. M., Schneider P., 2009, A&A, 499, 31
Hinshaw G. et al., 2013, ApJS, 208, 19
Kochanek C. S., 2002, ApJ, 578, 25

Kochanek C. S., Morgan N. D., Falco E. E., McLeod B. A., Winn J. N., Dembicky J., Ketzeback B., 2006, ApJ, 640, 47
Lesgourgues J., Tram T., 2014, J. Cosmol. Astropart. Phys., 9, 032
Lewis A., Bridle S., 2002, Phys. Rev. D, 66, 103511
Liao K. et al., 2015, ApJ, 800, 11
Linder E. V., 2011, Phys. Rev. D, 84, 123529
McCully C., Keeton C. R., Wong K. C., Zabludoff A. I., 2014, MNRAS, 443, 3631
McCully C., Keeton C. R., Wong K. C., Zabludoff A. I., 2016, preprint (arXiv:1601.05417)
Magain P., Courbin F., Sohy S., 1998, ApJ, 494, 472
Marshall P., Rajguru N., Slosar A., 2006, Phys. Rev. D, 73, 067302
Molinari N., Durand J., Sabatier R., 2004, Comput. Stat. Data Anal., 45, 159
Momcheva I. G., Williams K. A., Cool R. J., Keeton C. R., Zabludoff A. I., 2015, ApJS, 219, 29
Morgan N. D., Kochanek C. S., Pevunova O., Schechter P. L., 2005, AJ, 129, 2531
Oguri M., Marshall P. J., 2010, MNRAS, 405, 2579
Padmanabhan N., Xu X., Eisenstein D. J., Scalzo R., Cuesta A. J., Mehta K. T., Kazin E., 2012, MNRAS, 427, 2132
Pelt J., Kayser R., Refsdal S., Schramm T., 1996, A&A, 305, 97
Percival W. J. et al., 2010, MNRAS, 401, 2148
Planck Collaboration XVI, 2014, A&A, 571, A16
Planck Collaboration XIII, 2016a, A&A, 594, A13
Planck Collaboration XV, 2016b, A&A, 594, A15
Planck Collaboration XVIII, 2016c, A&A, 594, A18
Rathna Kumar S. et al., 2013, A&A, 557, A44
Refsdal S., 1964, MNRAS, 128, 307
Riess A. G. et al., 2016, ApJ, 826, 56
Rigault M. et al., 2015, ApJ, 802, 20
Salvatelli V., Marchini A., Lopez-Honorez L., Mena O., 2013, Phys. Rev. D, 88, 023531
Schneider P., Sluse D., 2013, A&A, 559, A37
Schneider P., Sluse D., 2014, A&A, 564, A103
Sluse D., Tewes M., 2014, A&A, 571, A60
Sluse D., Hutsemékers D., Courbin F., Meylan G., Wambsganss J., 2012, A&A, 544, A62
Sorce J. G., Tully R. B., Courtois H. M., 2012, ApJ, 758, L12
Spergel D. N., Flauger R., Hložek R., 2015, Phys. Rev. D, 91, 023518
Suyu S. H., Halkola A., 2010, A&A, 524, A94
Suyu S. H., Marshall P. J., Hobson M. P., Blandford R. D., 2006, MNRAS, 371, 983
Suyu S. H., Marshall P. J., Auger M. W., Hilbert S., Blandford R. D., Koopmans L. V. E., Fassnacht C. D., Treu T., 2010, ApJ, 711, 201
Suyu S. H. et al., 2012, ApJ, 750, 10
Suyu S. H. et al., 2013, ApJ, 766, 70
Suyu S. H. et al., 2014, ApJ, 788, L35
Tewes M., Courbin F., Meylan G., 2013a, A&A, 553, A120
Tewes M. et al., 2013b, A&A, 556, A22
Treu T., 2010, ARA&A, 48, 87
Treu T., Marshall P. J., 2016, A&AR, in press
Unruh S., Schneider P., Sluse D., 2016, preprint (arXiv:1606.04321)
Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, Phys. Rep., 530, 87
Wisotzki L., Christlieb N., Bade N., Beckmann V., Köhler T., Vanelle C., Reimers D., 2000, A&A, 358, 77
Wisotzki L., Schechter P. L., Bradt H. V., Heinmüller J., Reimers D., 2002, A&A, 395, 17
Wong K. C., Keeton C. R., Williams K. A., Momcheva I. G., Zabludoff A. I., 2011, ApJ, 726, 84
Wucknitz O., 2002, MNRAS, 332, 951
Xu D., Sluse D., Schneider P., Springel V., Vogelsberger M., Nelson D., Hernquist L., 2016, MNRAS, 456, 739

[1] *Laboratory of Astrophysics, Institute of Physics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, CH-1290 Versoix, Switzerland*
[2] *Max Planck Institute for Astrophysics, Karl-Schwarzschild-Strasse 1, D-85740 Garching, Germany*
[3] *Physik-Department, Technische Universität München, James-Franck-Straße 1, D-85748 Garching, Germany*
[4] *Institute of Astronomy and Astrophysics, Academia Sinica, PO Box 23-141, Taipei 10617, Taiwan*
[5] *Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 452 Lomita Mall, Stanford, CA 94035, USA*
[6] *Department of Physics, University of California, Davis, CA 95616, USA*
[7] *STAR Institute, Quartier Agora - Allée du six Août, 19c B-4000 Liège, Belgium*
[8] *Argelander-Institut für Astronomie, Auf dem Hügel 71, D-53121 Bonn, Germany*
[9] *National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan*
[10] *Institute of Cosmology and Gravitation, University of Portsmouth, Burnaby Rd, Portsmouth PO1 3FX, UK*
[11] *Department of Physics and Astronomy, University of California, Los Angeles, CA 90095, USA*
[12] *Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[13] *Exzellenzcluster Universe, Boltzmannstr. 2, D-85748 Garching, Germany*
[14] *Ludwig-Maximilians-Universität, Universitäts-Sternwarte, Scheinerstr. 1, D-81679 München, Germany*
[15] *Kapteyn Astronomical Institute, University of Groningen, PO Box 800, NL-9700-AV Groningen, The Netherlands*
[16] *Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan*
[17] *Physics Department, University of California, Santa Barbara, CA 93106, USA*

This paper has been typeset from a TEX/LATEX file prepared by the author.

4.3 Going further

In the light of the recent developments in precision cosmology, it seems likely that the current around H_0 could encompass more than a simple disagreement caused by unknown sources of error; the potential existence of new physics beyond the standard cosmological model is at stake. Obviously, such a leap forward cannot be taken without compelling evidence in its favor, which is currently far from being the case. Weinberg et al. (2013) showed that an independent determination of the Hubble constant with a 1% precision would drastically improve the precision with which other cosmological parameters can be constrained, likely representing one of our best observational probes to test our current knowledge of the Universe.

Reaching such a precision in H_0 will require many more strong lenses suitable to yield individual time-delay distance measurement at $\sim 5\%$ precision and below. The first step towards it is thus to find these lenses. A great wealth of data is currently being acquired by many ongoing wide-field surveys such as the Dark Energy Survey (DES, Dark Energy Survey Collaboration et al., 2016; Diehl, 2017), the Kilo-Degree Survey (KiDS, de Jong et al., 2013), the Hyper Suprime-Cam Survey (HSC, Chan et al., 2016) or the Canada-France Imaging Survey (CFIS), among others. These survey are completed by numerous lens-finding strategies, whether purely automated (a research topic currently flying high, see e.g. Joseph et al., 2014; Agnello et al., 2015; Chan et al., 2015; Avestruz et al., 2017; Lanusse et al., 2017; Petrillo et al., 2017; Schaefer et al., 2017; Sonnenfeld et al., 2017) or based on visual/spectroscopic inspection of the data (e.g. Marshall et al., 2016; More et al., 2016b,a). For example, the STRIDES program[6] included into the DES collaboration has already found around 30 new lensed quasars (A. Agnello, private communication) and an HST imaging follow-up proposal for 13 quads among them has been recently accepted. Figure 4.8 presents cut-outs of these candidates taken with ground based telescopes.

The second step is to improve the individual precision to which each system is analyzed. This requires sharper overall measurements, whether it be time-delay determination or high-resolution imaging and spectroscopy of the lens galaxy and its line-of-sight, but also improvements in the formalism and algorithms used to measure the time-delay distance to each lens from these observations. Observational requirements are currently being investigated (e.g. Meng et al., 2015; Linder, 2015) and current methods are developed further (Jee et al., 2016; McCully et al., 2017, and Sec. 3.2.4 of this manuscript). These efforts will be supported by the launch of the James Webb Space Telescope (JWST) in 2018 as well as the start of the operations of the Euclid mission in 2021 and the Large Synoptic Survey Telescope (LSST) in 2022. With all of this in mind, Treu and Marshall (2016) propose a roadmap for the upcoming decade, concluding on the practicability of a $\sim 1\%$ precision determination of the ensemble time-delay distance between 2020 and 2025.

In terms of cosmological output, the golden lens systems take the form of lensed type-Ia Supernovae. They are the currently brightest standard candles known, which makes them

---

[6] See http://strides.astro.ucla.edu.

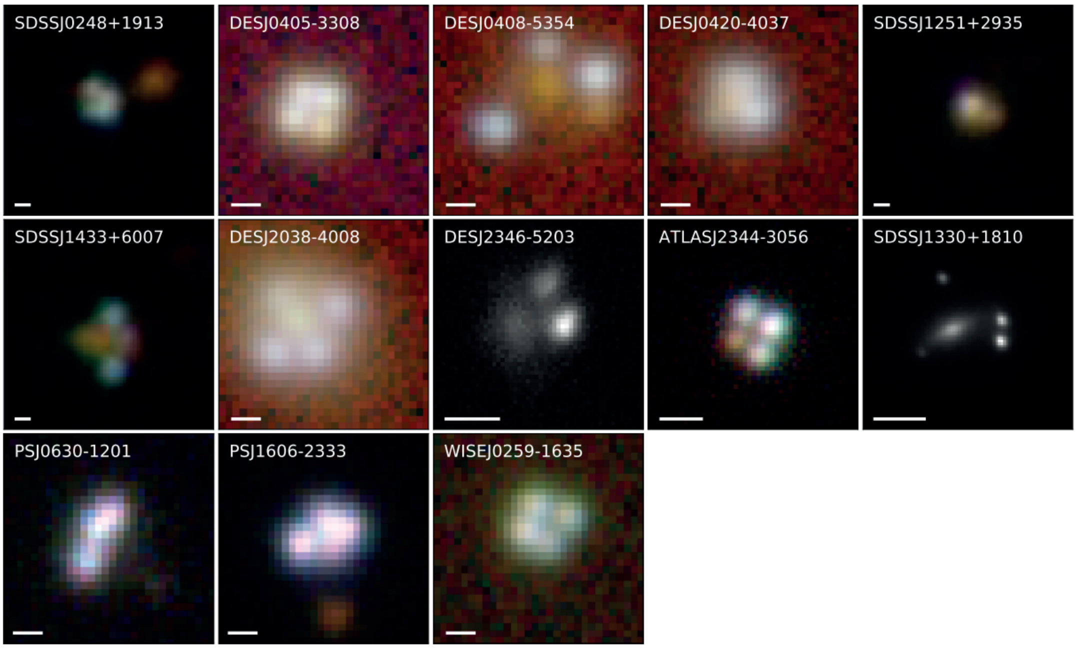


Figure 4.8: The sample of 13 quads planned for observations with the Hubble Space Telescope. False color images are extracted from either the SDSS, DES or PanSTARRS public image viewers. Exceptions are for DESJ2346, SDSSJ1330 and ATLASJ2344 whose cut-out images are from GEMINI GMOS imaging, Subaru AO K-band imaging and Magellan IMACS imaging, respectively. The bar on each image indicates the scale of one arcsecond. Courtesy of A. Agnello.

detectable by the current generation of instrument even when located at cosmological distances. They are, however, much more rare than lensed quasars: Oguri and Marshall (2010) predict a ratio of ~1 lensed Supernovae for 60 strong lenses, specifically for the LSST. So far, only two lensed Supernovae with resolved lensed images have been observed. The first one has been reported by Kelly et al. (2015) as a core-collapse Supernovae located in a spiral galaxy, lensed by a large foreground cluster. Mass modeling of the cluster predicted its reappearance a few years after its discovery (Diego et al., 2016; Oguri, 2015; Treu et al., 2016), which has been indeed later observed by Kelly et al. (2016) who forecasted a time-delay determination precise at a few percents. As stated in Treu et al. (2016), it was the only occasion known to date to perform a blind test of the lens modeling algorithms, especially on a very complex system such as a galaxy cluster. The drawback of this complexity is that it is very difficult to yield precise constraints on the lens modeling without any cosmological assumptions, which makes this system not suitable for time-delay cosmography. The second lensed Supernova with spatially resolved observations is, on the other hand, much more interesting in that regard. Goobar et al. (2017) reported the discovery of a type-Ia Supernova strongly lensed by a single foreground galaxy, along with a dozen of monitoring epochs around the maximum luminosity peak of the Supernova images. If type-Ia Supernovae indeed follow a known explosion pattern, this leads to two key advantages over quasars. First, the light curves of the images can be modeled following a template, which makes a precise time-delay determination possible in a couple of months of monitoring only. Second, the *absolute* luminosity of the source is known, which allows to measure the absolute magnification of the lensed images from the lens potential and thus breaks the mass-sheet degeneracy. However, using such systems for time-delay cosmography requires being able to monitor them at the time the Supernova explodes, notably in order to minimize the effect of chromatic microlensing on the observed light curves (Goldstein et al., 2017). Frequent observations of the sky in order to detect potential candidates early on are thus mandatory. The All-Sky Automated Survey for Supernovae (ASAN-SN, Kochanek et al., 2017) is a very interesting first step in that regard. Finally, let us note that in the great wealth of future lens systems to be discovered, there is a non zero probability of observing a "jackpot" lens, i.e. a type-Ia Supernova lensed by two galaxies at different redshifts. The two Einstein rings thus formed, along with the known absolute luminosity of the lensed images and the usual constraints on the mass profile could lift all the major sources of uncertainty and hopefully provide a nearly percent-level precision for H_0 on that single lens system.

In a closer future, the first rung of the H0LiCOW program will be completed with the study of WFI2033-4723 and HE1104-1805. The high-quality follow-up observations being already acquired, the cosmological constraints from these two targets are planned to be released by the end of 2017 and 2018, respectively. Suyu et al. (2016) - reproduced below - give an overview of the H0LiCOW program, notably presenting all the required observational efforts in order to properly process the five lenses of the first rung. For the two lenses still to be analyzed, and expecting an overall precision on the individual time-delay distance measurements similar to the three lenses already analyzed in H0LiCOW, Suyu et al. (2016) forecast a precision on H_0 below 3% in a flat-ΛCDM and below 3.5% in other cosmological models. In addition to this, the next H0LiCOW rung is already planned, with high-quality data currently being acquired and analyzed (or soon to be) for four new lenses.

Apart from the data acquisition, the main challenge in modern time-delay cosmography is the human investment it requires. To build on the example of HE0435-1223, carefully and thoroughly analyzing it took several months of intense work to dozens of experts in their own respective field. As pointed out by Treu and Marshall (2016), this raises the question of the scalability of the current analysis process; as long as the number of suitable strong lenses can be counted on the fingers of both hands (and feet), the cost will stay manageable. However, with the exponentially growing number of strong lenses candidate expected at the horizon 2025, the current analysis technique will need to be seriously adapted, certainly at the expense of the precision on individual systems. Whether high-precision cosmology from time-delay cosmography will be achieved through the statistical analysis and combination of thousands of lenses or by cherry-picking only the best targets is still unknown. The path to chose will strongly depend on the crucial work that is going to be conducted in the upcoming years by all the parties involved.

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **468,** 2590–2604 (2017)
Advance Access publication 2017 February 27

doi:10.1093/mnras/stx483

H0LiCOW – I. H_0 Lenses in COSMOGRAIL's Wellspring: program overview

S. H. Suyu,[1,2,3]★ V. Bonvin,[4] F. Courbin,[4] C. D. Fassnacht,[5] C. E. Rusu,[5] D. Sluse,[6] T. Treu,[7] K. C. Wong,[2,8] M. W. Auger,[9] X. Ding,[7,10] S. Hilbert,[11,12] P. J. Marshall,[13] N. Rumbaugh,[5] A. Sonnenfeld,[7,14,15] M. Tewes,[16] O. Tihhonova,[4] A. Agnello,[17] R. D. Blandford,[13] G. C.-F. Chen,[2,5] T. Collett,[18] L. V. E. Koopmans,[19] K. Liao,[7] G. Meylan[4] and C. Spiniello[1]

Affiliations are listed at the end of the paper

Accepted 2017 February 22. Received 2017 January 6; in original form 2016 June 29

ABSTRACT

Strong gravitational lens systems with time delays between the multiple images allow measurements of time-delay distances, which are primarily sensitive to the Hubble constant that is key to probing dark energy, neutrino physics and the spatial curvature of the Universe, as well as discovering new physics. We present H0LiCOW (H_0 Lenses in COSMOGRAIL's Wellspring), a program that aims to measure H_0 with <3.5 per cent uncertainty from five lens systems (B1608+656, RXJ1131−1231, HE 0435−1223, WFI2033−4723 and HE 1104−1805). We have been acquiring (1) time delays through COSMOGRAIL and Very Large Array monitoring, (2) high-resolution *Hubble Space Telescope* imaging for the lens mass modelling, (3) wide-field imaging and spectroscopy to characterize the lens environment and (4) moderate-resolution spectroscopy to obtain the stellar velocity dispersion of the lenses for mass modelling. In cosmological models with one-parameter extension to flat Λ cold dark matter, we expect to measure H_0 to <3.5 per cent in most models, spatial curvature Ω_k to 0.004, w to 0.14 and the effective number of neutrino species to 0.2 (1σ uncertainties) when combined with current cosmic microwave background (CMB) experiments. These are, respectively, a factor of ∼15, ∼2 and ∼1.5 tighter than CMB alone. Our data set will further enable us to study the stellar initial mass function of the lens galaxies, and the co-evolution of supermassive black holes and their host galaxies. This program will provide a foundation for extracting cosmological distances from the hundreds of time-delay lenses that are expected to be discovered in current and future surveys.

Key words: gravitational lensing: strong – quasars: individual: B1608+656, RXJ1131−1231, HE 0435−1223, WFI2033−4723, HE 1104−1805 – galaxies: structure – cosmological parameters – distance scale.

1 INTRODUCTION

In the past decade, the so-called flat Λ cold dark matter (ΛCDM) cosmological model consisting of dark energy (with density characterized by a cosmological constant Λ) and CDM in a spatially flat Universe has emerged as the standard cosmological model. This simple model has provided excellent fit to various cosmological observations including the temperature anisotropies in the cosmic microwave background (CMB) and galaxy density correlations in baryon acoustic oscillations (BAO). Recent CMB experiments, particularly the *Wilkinson Microwave Anisotropy Probe* (*WMAP*; Komatsu et al. 2011; Hinshaw et al. 2013) and the *Planck* satellite (Planck Collaboration XVI 2014; Planck Collaboration XIII 2016), and BAO surveys (e.g. Anderson et al. 2014; Kazin et al. 2014; Ross et al. 2015), have yielded stringent constraints with unprecedented precision on cosmological parameters in the spatially flat ΛCDM model.

An interesting result from *Planck* is its predicted value of the Hubble constant (H_0), a key cosmological parameter that sets the present-day expansion rate as well as the age, size and critical density of the Universe. *Planck* does not directly measure H_0, but rather enables its indirect inference through measurements of combinations of cosmological parameters given assumptions of the

---

★E-mail: suyu@mpa-garching.mpg.de

Published by Oxford University Press on behalf of the Royal Astronomical Society

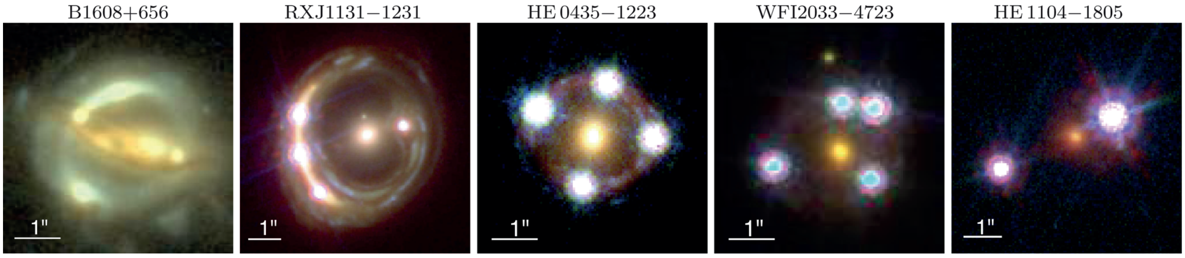


Figure 1. H0LiCOW lens sample, consisting of four quadruply lensed quasar systems in various configurations and one doubly lensed quasar system. The lens name is indicated above each panel. The colour images are composed using two (for B1608+656) or three (for other lenses) *HST* imaging bands in the optical and near-infrared. North is up and east is left.

background cosmological model. Intriguingly, *Planck*'s value of $H_0 = 67.8 \pm 0.9\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Planck Collaboration XIII 2016), from *Planck* temperature data and *Planck* lensing under the flat ΛCDM model, is lower than recent direct measurements based on the distance ladder, of $73.24 \pm 1.74\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ from the SH0ES program (Riess et al. 2016) and of $74.3 \pm 2.1\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Freedman et al. 2012) from the Carnegie-Chicago Hubble Program (Beaton et al. 2016). On the other hand, *Planck*'s H_0 value is similar to the results of some of the megamaser measurements (e.g. $H_0 = 68.9 \pm 7.1\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ from Reid et al. 2013, $H_0 = 73^{+26}_{-22}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ from Kuo et al. 2015 and $H_0 = 66.0 \pm 6.0\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ from Gao et al. 2016), although the uncertainties of these maser H_0 measurements are still substantial relative to that of *Planck*. A 1 per cent *direct* measurement of the Hubble constant is highly needed: such 1 per cent measurements of H_0 would address the possible tension with the CMB value which, if significant, would point towards deviations from the standard flat ΛCDM and new physics. In fact, when one relaxes, for example, the flatness or Λ assumption in the CMB analysis, strong parameter degeneracies between H_0 and other cosmological parameters appear, and the degenerate H_0 values from the CMB become compatible with the local H_0 measurements from the distance ladder (Freedman et al. 2012; Planck Collaboration XIII 2016; Riess et al. 2016). Thus, a 1 per cent measurement of H_0 is crucial for understanding the nature of dark energy, neutrino physics, the spatial curvature of the Universe and the validity of General Relativity (e.g. Hu 2005; Suyu et al. 2012a; Weinberg et al. 2013). In particular, the dark energy figure of merit of any survey that does not directly measure H_0 improves by ~40 per cent if H_0 is known to 1 per cent. Furthermore, independent methods to measure H_0 are necessary to overcome systematic effects, such as the known unknowns (e.g. the effects of crowding or metallicity dependence in the cosmic distance ladder) and the unknown unknowns in order to robustly verify or rule out the standard cosmological paradigm.

Strong gravitational lenses with measured time delays between the multiple images provide a competitive approach to measuring the Hubble constant, completely independent of the local distance ladder: we have demonstrated that we can constrain H_0 to ~7–8 per cent precision from a single time-delay lens system with ancillary data (Suyu et al. 2010, 2014). The time-delay method was first proposed by Refsdal (1964) even before the discovery of the first strong gravitational lens system (Walsh, Carswell & Weymann 1979), consisting of a foreground mass distribution that is located close along the line of sight to a background source (see Treu & Marshall 2016, for a recent review). The light from the background source is deflected by the foreground 'lens' mass distribution; such light bending produces distorted and, in rare cases of 'strong lensing', multiple and often spectacular images of the background source (e.g. Fig. 1).

When the background source is one that varies in its luminosity, such as an active galactic nucleus (AGN; e.g. Vanderriest et al. 1989; Schechter et al. 1997; Fassnacht et al. 1999, 2002; Kochanek et al. 2006; Courbin et al. 2011) or a supernova (SN; e.g. Quimby et al. 2014; Kelly et al. 2015, 2016; Goobar et al. 2016; Grillo et al. 2016; Kawamata et al. 2016; More et al. 2016b; Treu et al. 2016), the variability is manifest in each of the multiple images, but delayed in time relative to each other due to the different light paths. This time delay (Δt) thus depends on the 'time-delay distance' ($D_{\Delta t}$) and the lens mass distribution. Specifically, $\Delta t = D_{\Delta t} \Delta\phi / c$, where $\Delta\phi$ is the Fermat potential difference that is determined by the lens mass distribution and c is the speed of light. Therefore, by measuring the time delay from photometric light curves of the quasar images and modelling the lens mass distribution, one can determine the time-delay distance to the lens system and use the distance–redshift relation to constrain cosmological models.

More precisely, the time-delay distance is

$$D_{\Delta t} \equiv (1 + z_{\mathrm{d}}) \frac{D_{\mathrm{d}} D_{\mathrm{s}}}{D_{\mathrm{ds}}} \tag{1}$$

(Refsdal 1964; Suyu et al. 2010), where z_{d} is the redshift of the foreground deflector (also referred to as the strong lens), D_{d} is the angular diameter distance to the deflector, D_{s} is the angular diameter distance to the source and D_{ds} is the angular diameter distance between the deflector and the source. This time-delay distance is for a single strong-lens plane, with other line-of-sight mass distributions only weakly perturbing the strong-lens system and characterized via external shear and convergence. For cases where there are massive line-of-sight mass distributions at a different redshift from the strong-lens galaxy yet close in projection to it such that these massive structures cannot be well approximated by an external shear/convergence, it is necessary to use the multiplane lensing formalism (e.g. Blandford & Narayan 1986; Schneider, Ehlers & Falco 1992). In general, multilens plane ray tracing does not yield a single time-delay distance but rather several combinations of distances. None the less, even in some of these cases, we can derive an effective time-delay distance.

As a result of the unique combination of these three angular diameter distances, the time-delay distance $D_{\Delta t}$ is primarily sensitive to the Hubble constant, in contrast to other non-local distance probes such as SN that probe relative luminosity distances (e.g. Riess et al. 1998; Perlmutter et al. 1999; Conley et al. 2011;

Suzuki et al. 2012; Betoule et al. 2014) and BAO (e.g. Eisenstein et al. 2005; Percival et al. 2010; Blake et al. 2011; Anderson et al. 2014) that yield absolute angular diameter distances. We note though that BAO, together with the CMB, can be used to calibrate the absolute magnitude of SN; assuming that the absolute magnitude of SN does not evolve with redshift, this combination of BAO and SN provides an 'inverse-distance ladder' for the Hubble constant that is insensitive to assumptions on dark energy properties and spatial curvature (e.g. Heavens, Jimenez & Verde 2014; Aubourg et al. 2015). While BAO and the time-delay method both provide angular diameter distance measurements, the distinction is that BAO gives angular diameter distances at specific redshifts whereas the time-delay method yields time-delay distances ($D_{\Delta t}$) which are each a combination of three angular diameter distances. One could in fact determine the angular diameter distance to the lens $D_{\rm d}$ in addition to $D_{\Delta t}$ for time-delay lenses that have stellar velocity dispersion measurements of the foreground lens galaxy (Paraficz & Hjorth 2009; Jee, Komatsu & Suyu 2015). Without time delays, lenses with stellar velocity dispersion measurements can still offer a way to determine the cosmological matter and dark energy density parameters via a ratio of angular diameter distances (e.g. Futamase & Hamana 1999; Futamase & Yoshida 2001; Grillo, Lombardi & Bertin 2008). Recently, Jee et al. (2016) have shown that measurements of $D_{\Delta t}$ and $D_{\rm d}$ from a modest sample of time-delay lenses with lens velocity dispersion measurements yield competitive constraints on cosmological models. In practice, both distances appear as intermediate quantities between the sought after cosmological parameters and the observed quantities.

In order to measure distances precisely and accurately from time-delay lenses, we need four key ingredients in addition to the spectroscopic redshifts of the lens and the source: (1) time delays, (2) high-resolution and high signal-to-noise ratio images of the lens systems, (3) characterization of the lens environment and (4) stellar velocity dispersion of the lens galaxy. These can be obtained via imaging and spectroscopy from *Hubble Space Telescope* (*HST*) and ground-based observatories. In Section 2, we detail each of these requirements.

We initiated the H0LiCOW (H_0 Lenses in COSMOGRAIL's Wellspring) program with the aim of measuring the Hubble constant with better than 3.5 per cent precision and accuracy (in most background cosmological models), through a sample of five time-delay lenses. We obtain the key ingredients to each of the lenses through observational follow-ups and novel analysis techniques. In particular, we have high-quality lensed quasar light curves, primarily obtained via optical monitoring by the COSMOGRAIL (COSmological MOnitoring of GRAvItational Lenses; e.g. Courbin et al. 2005; Vuissoz et al. 2008; Courbin et al. 2011; Tewes et al. 2013b) and Kochanek et al. (2006) teams but also via radio-wavelength monitoring (Fassnacht et al. 2002). COSMOGRAIL has been monitoring more than 20 lensed quasars for more than a decade. The unprecedented quality of the light curves combined with new curve-shifting algorithms (Tewes, Courbin & Meylan 2013a) lead to time delays with typically ∼3 per cent accuracy (Fassnacht et al. 2002; Courbin et al. 2011; Tewes et al. 2013b). In addition, we obtain *HST* imaging that reveal the 'Einstein ring' of the lens systems in high resolution, and develop state-of-the-art lens modelling techniques (Suyu et al. 2009; Suyu & Halkola 2010; Suyu et al. 2012b) and kinematic modelling methods (Auger et al. 2010; Sonnenfeld et al. 2012) to obtain the lens mass distribution with a few percent uncertainty (e.g. Suyu et al. 2013, 2014). We further obtain wide-field imaging and spectroscopy to characterize the environment of the field, as well as the spectroscopy of the lens galaxy to obtain the stellar velocity dispersion. The exquisite follow-up data set that we have acquired allow us not only to constrain cosmology but also to study lens galaxy and source properties for understanding galaxy evolution, including the dark matter distribution in galaxies, the stellar initial mass function of galaxies and the co-evolution between supermassive black holes and their host galaxies.

A crucial aspect of our program is the use of blind analysis (e.g. Conley et al. 2006; Suzuki et al. 2012; Suyu et al. 2013; von der Linden et al. 2014) to test for residual systematics and avoid subconscious experimenter bias. In particular, we have developed core analysis techniques for the first lens whose dissection was not blinded (B1608+656; Suyu et al. 2010); we subsequently build upon these techniques and perform blind analysis on the other lenses in the sample. In the blind analysis, the idea is not to blind all the model parameters being inferred, but rather just the cosmological parameters that we aim to measure (as well as any derived parameters or summary statistics from which we could infer the cosmological parameters). We therefore blind the time-delay distance and all cosmological parameters in our analysis. Specifically, throughout the analysis, we only ever plot these blinded parameters offset by their posterior median value. We can then still use the parameter correlations and the uncertainties to cross check our analysis, since the temptation to stop investigating systematic errors when the 'right answer' has been obtained has been removed. Only when the collaboration deems the analysis to be final and complete do we 'open the box' to reveal the median values of the parameters, and then publish these results without modifications.

This paper (hereafter H0LiCOW Paper I) is the first of the series, and gives an overview of the program. There are four more papers that detail the data sets and analysis of the H0LiCOW lens system HE 0435−1223. In particular, Sluse et al. (2017, hereafter H0LiCOW Paper II) present the spectroscopic follow-up of the strong-lens field to measure redshifts of massive and nearby objects close in projection to the strong-lens system and identify galaxy groups along the line of sight. Rusu et al. (2017, hereafter H0LiCOW Paper III) use our multiband wide-field imaging to characterize the lens environment in combination with ray tracing with numerical simulations. Wong et al. (2017, hereafter H0LiCOW Paper IV) perform the lens mass modelling of the strong-lens system incorporating the time delays, high-resolution imaging and lens stellar kinematics data sets to infer the distance to the lens via blind analysis. Bonvin et al. (2017, hereafter H0LiCOW Paper V) present the time-delay measurements from COSMOGRAIL lens monitoring and the cosmological inference based on the previous three papers.

The outline of this paper is as follows. We describe the key ingredients for time-delay cosmography in Section 2, present the five H0LiCOW lens systems in Section 3 and describe our observational campaign in Section 4. The key components of the four analysis papers introduced above are summarized in Section 5. We show the forecasted cosmographic constraints from the H0LiCOW sample in Section 6. We summarize in Section 7 with an outlook for the program.

2 OBSERVATIONAL REQUIREMENTS OF THE TIME-DELAY METHOD

In this section, we describe the observational requirements of the four ingredients for accurate and precise distance measurements from time-delay lenses.

(i) *Time delays*. Monitoring campaigns to map out the variability of the multiple lensed images over time have been carried out both in the radio and optical wavelengths (e.g. Vanderriest et al. 1989; Schechter et al. 1997; Burud et al. 2002; Fassnacht et al. 2002; Hjorth et al. 2002; Kochanek et al. 2006; Vuissoz et al. 2007; Rumbaugh et al. 2015). Regular and frequent observations, at least once every few days, are necessary so that the variability pattern of the background source can be observed in each of the multiple images and be matched up to obtain the time delays. Monitoring in the optical requires a long baseline or high photometric precision to overcome systematic variations due to microlensing by stars in the lensing galaxy that could be mistaken as the background source intrinsic variability (e.g. Tewes et al. 2013b; Sluse & Tewes 2014). Curve-shifting methods have been developed to measure the time delays from the light curves (e.g. Press, Rybicki & Hewitt 1992; Pelt et al. 1996; Fassnacht et al. 2002; Harva & Raychaudhury 2008; Morgan et al. 2008; Hirv, Olspert & Pelt 2011; Hojjati, Kim & Linder 2013; Tewes, Courbin & Meylan 2013a). A recent time-delay challenge showed that some of the methods can recover accurately the time delays in a blind test (Dobler et al. 2015; Liao et al. 2015), particularly the methods we use from the COSMOGRAIL collaboration (e.g. Tewes et al. 2013a; Bonvin et al. 2016).

(ii) *Well-resolved lensed images*. The strong-lensing information, such as the multiple image positions of the background source, is needed to obtain the foreground lens mass distribution for converting the time delays into distances. Deep and high-resolution imaging of the strong-lens system reveal the 'Einstein rings' that are the spatially extended and lensed images of the background source, such as the host galaxy of the AGN. In the past decade, methods have been developed to take advantage of the thousands of intensity pixels of the extended images to constrain precisely within a few percent the lens potential at the location of the multiple images (e.g. Kochanek, Keeton & McLeod 2001; Warren & Dye 2003; Treu & Koopmans 2004; Koopmans 2005; Dye et al. 2008; Suyu et al. 2009; Vegetti & Koopmans 2009; Suyu et al. 2013; Birrer, Amara & Refregier 2015; Chen et al. 2016). The time-delay distance is particularly sensitive to the radial profile of the lens galaxy mass distribution (e.g. Kochanek 2002; Wucknitz 2002; Wucknitz, Biggs & Browne 2004; Suyu 2012). Imaging with high-signal-to-noise ratio and high angular resolution of the Einstein ring helps to constrain the lens radial profile in the region of the ring, and hence the time-delay distance, up to a mass-sheet transformation (described below).

(iii) *The lens environment*. The distribution of mass external to the lens galaxy, such as that associated with galaxies which are close in projection to the lens system along the line of sight, affects the time delays between the multiple images and hence our cosmological distance measurements. An external convergence $\kappa_{\rm ext}$ can be absorbed by the lens and source model leaving the fit to the lensed images unchanged, but the predicted time delays altered by a factor of $(1-\kappa_{\rm ext})$.

To break this 'mass-sheet degeneracy' (MSD; Falco, Gorenstein & Shapiro 1985), one can study the environment of the lens system to constrain $\kappa_{\rm ext}$ within a few percent[1] through spectroscopic/photometric observations of local galaxy groups and line-of-sight structures (e.g. Fassnacht et al. 2006; Momcheva et al. 2006, 2015) in combination with ray tracing through numerical N-body simulations (e.g. Hilbert et al. 2007, 2009; Suyu et al. 2010; Collett et al. 2013; Greene et al. 2013). Furthermore, McCully et al. (2014, 2016) developed a new framework to model line-of-sight mass distributions efficiently and quantified the environment effects through realistic simulations of lens fields. By reconstructing the three-dimensional mass distribution of strong-lens sightlines, McCully et al. (2016) can obtain constraints on $\kappa_{\rm ext}$ that are consistent with but tighter than those from the aforementioned statistical approach of combining galaxy number density observations with N-body simulations (see also Collett et al. 2013 whose sightline mass reconstruction also produces tighter constraints on $\kappa_{\rm ext}$ than the statistical approach). Recently, Collett & Cunnington (2016) have pointed out that the external convergence over an ensemble of lenses usually does not average to zero – lenses, like typical massive galaxies, preferentially live in locally overdense regions (Holder & Schechter 2003; Treu et al. 2009; Fassnacht, Koopmans & Wong 2011) and are therefore slightly easier to detect and monitor. None the less, this bias in detection and/or selection that is due to overdensity is expected to have currently negligible impact on $D_{\Delta t}$ (<1 per cent impact). In contrast, measurements of $D_{\rm d}$ that come from combining delays with the lens velocity dispersion are impervious to $\kappa_{\rm ext}$ (Jee et al. 2015).

(iv) *The lens galaxy stellar velocity dispersion*. The combination of lensing and stellar kinematics is a powerful probe of the lens galaxy mass distribution (e.g. Romanowsky & Kochanek 1999; Treu & Koopmans 2002; Koopmans et al. 2003; Barnabè et al. 2009, 2011; Sonnenfeld et al. 2012) since the combination breaks degeneracies that are inherent in each approach, and in particular the mass-sheet degeneracy in lensing. Schneider & Sluse (2013) pointed out that the mass-sheet degeneracy can manifest as a lens mass profile degeneracy, which Xu et al. (2016) investigated using simulated galaxies. Moreover, the mass-sheet degeneracy is in fact a special case of a more general 'source-position transformation' (Schneider & Sluse 2014; Unruh, Schneider & Sluse 2016), although this latter transformation typically does not leave the multiple time delays invariant. To break such lensing degeneracies, information from the lens galaxy stellar kinematics is crucial: Suyu et al. (2014) showed that the lens velocity dispersion substantially reduced the dependence of the time-delay distance on lens mass profile assumptions. The lens velocity dispersion is also a key ingredient for measuring $D_{\rm d}$, which is more sensitive to dark energy properties than $D_{\Delta t}$ (Jee et al. 2015, 2016).

3 H0LICOW SAMPLE OF LENSES

In Fig. 1, we show the images of the five lenses in our sample. The left four lenses are quadruply lensed quasar systems (quads) and the rightmost lens system is a doubly lensed quasar system (double). As described below, the four quads span the three generic multiple image configurations we have in galaxy-scale strong lenses: symmetric, fold (with two merging images) and cusp (with three merging images). Therefore, our sample will allow us to explore to some extent the optimal image configuration for cosmographic studies.

Our sample of lenses was chosen based on three criteria: (1) availability of accurate and precise time delays, (2) existing measurements of spectroscopic redshifts for both the lens and the background source and (3) the lens system is not located near a galaxy cluster (to avoid potentially large systematic effects due to mass along the line of sight). We prefer quads to doubles since quads provide more observational constraints on the mass model (e.g. more time delays and image positions). The four quads in our sample were the only known quad lenses that passed the above three criteria at the time of our sample selection. There were a few doubles

---

[1] In terms of its impact on $D_{\Delta t}$.

that pass these criteria, and we chose HE 1104−1805 as the first double in this pilot program given its relative simplicity for mass modelling with only one strong-lens galaxy (in contrast to other systems that have multiple massive lens galaxies). We describe in more detail each of the lenses below.

B1608+656. The lens system was discovered in the Cosmic Lens All-Sky Survey (Myers et al. 1995; Browne et al. 2003; Myers et al. 2003). The radio-loud AGN is lensed into four images that are relatively dim in the optical wavelength, thus showing clearly the extended Einstein ring of the AGN host galaxy in the *HST* imaging (Fig. 1). Two of the four multiple images are close together, making this a standard 'fold' configuration. The system contains two lens galaxies that appear to be interacting and resulting in dust extinction in the system (e.g. Koopmans et al. 2003; Surpi & Blandford 2003; Suyu et al. 2009). The lens and source redshifts are, respectively, $z_{\rm s} = 1.394$ (Fassnacht et al. 1996) and $z_{\rm d} = 0.6304$ (Myers et al. 1995). This system was the first quad lens with all three time delays measured with uncertainties of only a few percent (Fassnacht et al. 1999, 2002).

RXJ1131−1231. Sluse et al. (2003) discovered RXJ1131−1231 serendipitously during polarimetric imaging of a sample of radio quasars. This system shows a spectacular Einstein ring, with multiple arclets that are the lensed images of the AGN host galaxy containing a bulge and a disk with spiral arms and star formation clumps. Three of the four quasar images are close to each other, forming the typical 'cusp' configuration. The lens redshift is at $z_{\rm d} = 0.295$ (Sluse et al. 2003, 2007), and the source redshift is at $z_{\rm s} = 0.654$ (Sluse et al. 2007).[2]

HE 0435−1223. This lens system was found by Wisotzki et al. (2002), originally selected in the Hamburg/ESO survey (Wisotzki et al. 2000) as a highly probable quasar candidate. The background quasar is lensed into four multiple images that are nearly symmetrically positioned in the 'cross' configuration. The background source is at redshift $z_{\rm s} = 1.693$ (Sluse et al. 2012)[3] and the foreground strong lens is at redshift $z_{\rm d} = 0.4546$ (Morgan et al. 2005; Eigenbrod et al. 2006). The *HST* image reveals an elliptical ring that connects the four images of the AGN. This ring is produced by the extended lensed images of the AGN galaxy.

WFI2033−4723. Morgan et al. (2004) discovered this quad lens system as part of an optical imaging survey using the MPG/ESO 2.2-m telescope at La Silla, Chile that is operated by the European Southern Observatory (ESO). The lens system exhibits a typical fold configuration, since it contains two merging quasar images. The quasar is at redshift $z_{\rm s} = 1.662$ (Sluse et al. 2012), which is consistent with the first measurement by Morgan et al. (2004). The quasar images are substantially brighter than the background quasar host galaxy and the foreground lens galaxy. Morgan et al. (2004) identified the foreground lens galaxy, whose redshift was measured to be $z_{\rm d} = 0.661$ (Eigenbrod et al. 2006), consistent with an earlier measurement by Ofek et al. (2006). The high-resolution *HST* imaging shows several galaxies in the vicinity of the lens system. Since these galaxies would likely influence the lens potential, their redshifts will be obtained with our ancillary data (Section 4.3) in order to incorporate them into the lens mass model.

HE 1104−1805. This system was also discovered in the early phase of the Hamburg/ESO survey by Wisotzki et al. (1993). The two lensed quasar images are separated by $\sim 3''$ and is unusual in having the brighter image as the one closer to the foreground lens galaxy, which was first identified by Courbin, Lidman & Magain (1998) and Remy et al. (1998). The source is at $z_{\rm s} = 2.316$ (Smette et al. 1995), and the lens is at a relatively high redshift of $z_{\rm d} = 0.729$ (Lidman et al. 2000). The *HST* image shows multiple luminous structures/galaxies around the lens system.

4 OBSERVATIONAL FOLLOW-UP

In collaboration with the COSMOGRAIL team, we carry out an observational campaign in order to obtain each of the four ingredients for distance measurements of the H0LiCOW lenses. We describe the monitoring in Section 4.1 to get the time delays, deep *HST* imaging to constrain the lens galaxy mass distribution in Section 4.2, wide-field spectroscopy and imaging to study the lens environment in Section 4.3 and spectroscopy of the foreground lens galaxy to measure the stellar velocity dispersion in Section 4.4.

4.1 Time delays

Of the five H0LiCOW lenses, B1608+656 has been monitored previously by Fassnacht et al. (1999, 2002) using the Very Large Array, whereas the other four lenses are currently being monitored by the COSMOGRAIL and Kochanek et al. (2006) collaborations using a network of 1–2 m optical telescopes, particularly the Euler telescope in Chile.

Using three seasons of monitoring of B1608+656, especially the third season that showed significant variability that repeated in all four quasar images, Fassnacht et al. (2002) measured all three relative time delays between the four quasar images with uncertainties of a few percent. The image fluxes were measured every 3–4 d during the monitoring. The time delays span ∼30–80 d, relative to the first image that varies.

The monitoring of RXJ1131−1231, HE 0435−1223, WFI2033−4723 and HE 1104−1805 by the COSMOGRAIL and Kochanek et al. (2006) teams started in 2003, with a photometric point every 2–4 d. The MCS deconvolution method (Magain, Courbin & Sohy 1998; Cantale et al. 2016) is used to extract the photometry of the quasar images for building the light curves. Tewes et al. (2013a) set up an automated pipeline to reduce the images, build the light curves and measure the time delays using a state-of-the-art curve-shifting algorithm that simultaneously models both intrinsic variability of the AGNs and microlensing variations. With this pipeline, Bonvin et al. (2016) recovered the time delays with a precision of ∼3 per cent and negligible bias for simulated light curves mimicking COSMOGRAIL monitoring in the blind strong-lens time delay challenge (Liao et al. 2015), demonstrating the robustness of their curve-shifting algorithms.

The monitoring and analysis yield time delays in RXJ1131−1231 with a 1.5 per cent uncertainty on the longest delay (Tewes et al. 2013b). The light curve has been separately modelled by A. Hojjati and E. Linder using the Gaussian process technique (Hojjati et al. 2013), who have obtained delays that are consistent with the measurements of Tewes et al. (2013b) (Linder, private communication). The monitoring and analysis of HE 0435−1223 are described

---

[2] The source redshift of $z_{\rm s} = 0.654$ is based on the narrow emission lines, which is considered more accurate than the H α and Mg II lines (Hewett & Wild 2010) that yield $z_{\rm s} = 0.657$ (Sluse et al. 2007). We note that a 0.003 change in $z_{\rm s}$ corresponds to a <0.4 per cent change in $D_{\Delta t}$ for RXJ1131−1231, and even less change in $D_{\Delta t}$ for the other higher redshift lens systems.

[3] Based on Mg II emission line, which results in a slightly higher redshift value than the previous measurement of $z_{\rm s} = 1.689$ (Wisotzki et al. 2002) from C IV line that is known to be prone to systematic blueshifts in many quasars.

in H0LiCOW Paper V, with a relative uncertainty of 6.5 per cent on the longest delay (between images A and D). The measurement precision in the delays has improved by a factor of 2 compared to the previous measurements by Courbin et al. (2011) with the five additional years of monitoring and improvements in the curve-shifting algorithms. For WFI2033−4723 and HE 1104−1805, we expect to improve on the previous delay measurements by Vuissoz et al. (2008) and Poindexter et al. (2007), respectively, with the new curve-shifting techniques, and estimate relative uncertainties of $\sim$4 per cent and $\sim$2 per cent, respectively, from the monitoring campaign.

4.2 *HST* observations

Deep *HST* Advanced Camera for Surveys (ACS) observations of B1608+656 were obtained in Program 10158 (PI: C. D. Fassnacht) in two filters, *F*606*W* and *F*814*W*. Suyu et al. (2009) have described these observations in detail. Furthermore, Suyu et al. (2009) analysed these data and used a pixelated lens potential reconstruction technique to model the lens mass distribution, which were subsequently used for cosmographic analysis in Suyu et al. (2010).

Archival *HST* ACS observations of RXJ1131−1231 (Program 9744; PI: C. S. Kochanek) are available in two filters, *F*555*W* and *F*814*W*. Details of the observations are described in, e.g. Claeskens et al. (2006). These have been used to model the lens mass distribution for cosmography, accounting for uncertainties due to assumptions on the lens mass profile (Suyu et al. 2013, 2014). Recently, Birrer, Amara & Refregier (2016) have also used these observations to model independently the lens mass distribution of RXJ1131−1231 for cosmography, obtaining results that are consistent with Suyu et al. (2013).

We have obtained new deep *HST* Wide Field Camera 3 (WFC3) observations in Program 12889 (PI: S. H. Suyu) of the remaining three lenses (HE 0435−1223, WFI2033−4723 and HE 1104−1805) in the infrared (IR) channel. The goal of these observations is to detect the Einstein rings of the AGN host galaxies at high signal-to-noise ratios, in order to constrain the foreground lens mass distribution (previous *HST* observations had insufficient signal-to-noise ratios of the rings for our analysis). We use the *F*160*W* filter to optimize the contrast between the AGN host galaxy and the AGN, since the host galaxy is brighter in the IR compared to the optical, especially for HE 1104−1805 where the quasar is at a high redshift.

We employ four-point dither patterns that trace out parallelograms with the lengths of the sides being non-integral numbers of pixels. For each lens, we use multiple parallelograms that are offset by non-integral pixels. Specifically, we use 2, 5 and 3 parallelograms for HE 0435−1223, WFI2033−4723and HE 1104−1805, respectively, depending on the total exposure time needed to image the Einstein ring. We further ensure that the dithering points do not overlap to avoid IR persistence effects. This dithering strategy allows us to recover effectively an angular resolution of $\sim$0.08arcsec from the native 0.13 arcsec pixel scale.

Table 1. New *HST* WFC3/IR Observations of HE 0435−1223, WFI2033−4723 and HE 1104−1805.

Lens	Date	Number/type of exposures	Time (s) per exposure
HE 0435−1223	2012-10-28	8 short exp.	44
		15 long exp.[4]	599
WFI2033−4723	2013-05-03 to 2013-05-04	20 short exp.	74
		4 long exp.[4]	599
		32 long exp.[4]	699
HE 1104−1805	2013-03-18	12 short exp.	26
		24 long exp.	599

Notes. At each dither position, an exposure sequence of short–long–long exposure times is adopted in order to sample the large dynamical range of the AGN and its much fainter host galaxy.[4]

Since the AGN host galaxy is substantially fainter than the AGN, we further adopt an exposure sequence of short–long–long at each of the dithering point.[4] The first short exposure allows us to characterize the AGN, whereas the long exposures would get the AGN host with possibly the pixels near the bright AGN saturated. We note that there are multiple non-destructive reads during each exposure with the MULTIACCUM mode of the WFC3/IR detector, so we can have a count rate estimate on the AGN pixels even in the long exposures if several non-destructive reads are available before saturation. The short exposures are taken to ensure that there are sufficient reads to characterize accurately the pixel count rates near the AGN positions, in case the long exposures are indeed saturated with insufficient non-destructive reads. In essence, the combination of the short and long exposures allows us to reconstruct in full the brightness distribution of *both* the lensed AGN and the lensed host galaxy. We summarize our observations in Table 1.

We reduce the images using DRIZZLEPAC.[5] The images are drizzled to a final pixel scale of 0.08 arcsec, without masking the bright AGN pixels as they are well characterized by the short exposures. The uncertainty on the flux in each pixel is estimated from the science image and the drizzled exposure time map by adding in quadrature the Poisson noise from the source and the background noise due to the sky and detector readout.

In Fig. 2, we show the reduced *HST* WFC3 observations of HE 0435−1223, WFI2033−4723 and HE 0435−1223 in the top panels from left to right. In the bottom, we show the images with the lens light subtracted with GLEE,[6] revealing the Einstein ring of the AGN host galaxy. In H0LiCOW Paper IV, we detail the modelling of HE 0435−1223 using multilens-plane ray tracing (e.g. Blandford & Narayan 1986; Schneider et al. 1992; Blandford & Kochanek 2004) and point spread function (PSF) reconstruction techniques developed by Suyu et al. (in preparation). The subtraction of lens light in WFI2033−4723 and HE 1104−1805 (bottom-middle and bottom-right panels of Fig. 2, respectively) is based on an initial PSF built from stars in the field without any lens mass modelling or iterative PSF reconstruction, hence the lens-subtraction residuals. Furthermore, the lens galaxy of HE 1104−1805 is on a diffraction spike of the brighter AGN image – an accurate PSF model would be crucial for distinguishing the lens galaxy, the two AGN images and the lensed host galaxy of the AGN. The full modelling and

---

[4] For HE 0435−1223, one long exposure was lost due to a satellite passing over the target. For one of the parallelogram dither pattern for WFI2033−4723, we use an exposure sequence of short–long (rather than short–long–long) at each dither position to optimize target exposure time given overhead associated with observations.

[5] DRIZZLEPAC is a product of the Space Telescope Science Institute, which is operated by AURA for NASA.

[6] A lens modelling software package developed by A. Halkola and S. H. Suyu (Suyu & Halkola 2010; Suyu et al. 2012b).

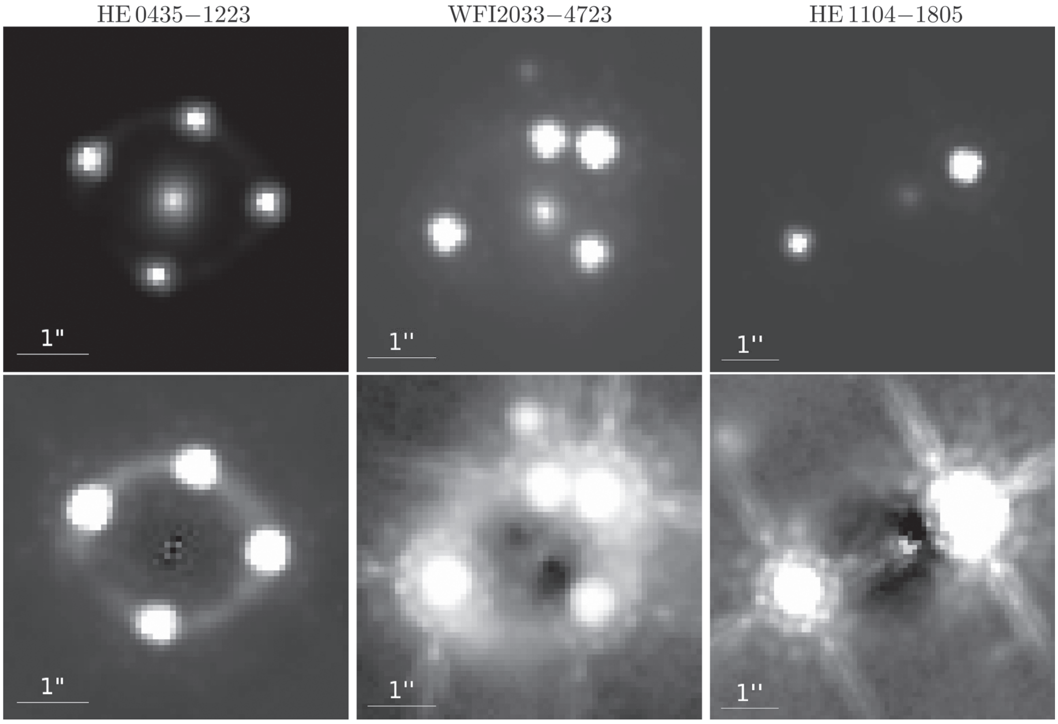


Figure 2. *HST* WFC3 *F*160*W* observation of HE 0435−1223, WFI2033−4723 and HE 1104−1805 from left to right in the top panels. In the bottom panels, the lens-galaxy light has been subtracted, revealing the Einstein ring of the AGN host galaxy that is needed for accurate and precise lens mass modelling. The full modelling of HE 0435−1223 is detailed in H0LiCOW Paper IV. The lens subtraction for WFI2033−4723 and HE 1104−1805 in the bottom-middle and bottom-right panels, respectively, is based on an initial PSF model without PSF reconstruction (which we defer to future work), hence the visible residuals. In each of the panels, north is up and east is left.

analysis of WFI2033−4723 and HE 1104−1805 will appear in future publications.

4.3 Wide-field spectroscopy and imaging of lens environment

We obtain wide-field spectroscopy to pinpoint the redshifts of the bright galaxies in the fields of the H0LiCOW lenses, particularly the ones close to the strong lens. Redshifts of nearby galaxies, especially those within a few arcseconds from the strong lens, are crucial since the external convergence approximation is often insufficient for these galaxies (e.g. McCully et al. 2014) and they need to be incorporated directly into the strong-lens modelling. We use the multiobject spectrographs on the Very Large Telescope, the Gemini Telescope and the W. M. Keck Telescope to target our lens fields, as summarized in Table 2. The spectroscopic redshifts and galaxy group identifications are detailed in Fassnacht et al. (2006), H0LiCOW Paper II, and forthcoming publications.

To further characterize the lens environment and determine κ_{ext}, we obtain wide-field multiband imaging using the Canada–France–Hawaii Telescope, Subaru Telescope, the Very Large Telescope, Gemini Telescope and *Spitzer Space Telescope*. Table 3 summarizes the follow-up imaging that allow us to compute the photometric redshifts of structures along the line of sight as well as to estimate their stellar masses. Details of the observations and inference on κ_{ext} are described in H0LiCOW Paper III and forthcoming publications.

Williams et al. (2006) have independently obtained *I* and either *V* or *R* images of all five H0LiCOW lenses using the 4-m Cerro Tololo Inter-American Observatory (CTIO) Blanco telescope for the southern fields and the 4-m Kitt Peak National Observatory Mayall telescope for the northern fields. Using these images to select spectroscopic targets, Momcheva et al. (2015) have obtained spectroscopic observations of the five H0LiCOW lenses using the

Table 2. Wide-field spectroscopy of H0LiCOW lenses as part of the H0LiCOW program.

Lens	Facility/instrument	Proposal PI
B1608+656	W. M. Keck/LRIS	C. D. Fassnacht
	W. M. Keck/ESI	C. D. Fassnacht
RXJ1131−1231	W. M. Keck/LRIS	C. D. Fassnacht
HE 0435−1223	W. M. Keck/LRIS	C. D. Fassnacht
	VLT/FORS2	D. Sluse
	Gemini/GMOS	T. Treu
WFI2033−4723	VLT/FORS2	D. Sluse
	Gemini/GMOS	T. Treu
HE 1104−1805	VLT/FORS2	D. Sluse
	Gemini/GMOS	T. Treu

Notes. Abbreviations are LRIS (Low-Resolution Imaging Spectrometer; Oke et al. 1995; Rockosi et al. 2010), ESI (Echellete Spectrograph and Imager; Sheinis et al. 2002), VLT (Very Large Telescope), FORS2 (FOcal Reducer and low dispersion Spectrograph; Appenzeller et al. 1998) and GMOS (Gemini Multi-Object Spectrographs; Hook et al. 2004). Details of the observations for B1608+656 are in Fassnacht et al. (2006), and for the other four lenses are in H0LiCOW Paper II and forthcoming publications. Additional integral field spectroscopy of the central 30 arcmin around WFI2033−4723 has been recently obtained with the Multi Unit Spectroscopic Explorer (MUSE; Bacon et al. 2012) on the VLT.

Table 3. Wide-field imaging obtained as part of the H0LiCOW program.

Lens	Facility/instrument	Wavelength bands	Proposal PI
B1608+656	CFHT/MegaCam	u	S. H. Suyu
	Subaru/Suprime-Cam	g, r, i	C. D. Fassnacht
	Subaru/MOIRCS	J, H, $K_{\rm s}$	C. D. Fassnacht
	Gemini/NIRI	J, $\mathrm{K_s}$	C. D. Fassnacht
	Spitzer/IRAC	3.6 μm, 4.5 μm	C. E. Rusu
RXJ1131−1231	CFHT/MegaCam	u	S. H. Suyu
	Subaru/Suprime-Cam	g, r, i	C. D. Fassnacht
	Subaru/MOIRCS	J, H, $\mathrm{K_s}$	C. D. Fassnacht
	Gemini/NIRI	J, $\mathrm{K_s}$	C. D. Fassnacht
HE 0435−1223	CFHT/MegaCam	u	S. H. Suyu
	Subaru/Suprime-Cam	g, r, i	C. D. Fassnacht
	Subaru/MOIRCS	H	C. D. Fassnacht
	Gemini/NIRI	J, $\mathrm{K_s}$	C. D. Fassnacht
WFI2033−4723	CTIO Blanco/DECam	u	C. E. Rusu
	VLT/HAWK-I	J, H, K	C. D. Fassnacht
HE 1104−1805	CFHT/MegaCam	u	S. H. Suyu
	Subaru/Suprime-Cam	g, r, i	C. D. Fassnacht
	Subaru/MOIRCS	J, H, $\mathrm{K_s}$	C. D. Fassnacht
	Gemini/NIRI	J, $\mathrm{K_s}$	C. D. Fassnacht

Notes. Abbreviations and references for the instruments are CFHT (Canada–France–Hawaii Telescope) MegaCam (Boulade et al. 2003), Suprime-Cam (Miyazaki et al. 2002), MOIRCS (Multi-Object InfraRed Camera and Spectrograph; Suzuki et al. 2008; Ichikawa et al. 2006), NIRI (Near InfraRed Imager and Spectrometer; Hodapp et al. 2003), IRAC (Infrared Array Camera; Fazio et al. 2004), CTIO (Cerro Tololo Inter-American Observatory) DECam (Dark Energy Camera; Diehl & Dark Energy Survey Collaboration 2012), VLT (Very Large Telescope) HAWK-I (High Acuity Wide field K-band Imager; Pirard et al. 2004; Casali et al. 2006; Kissler-Patig et al. 2008). Details of the observations are in H0LiCOW Paper III and forthcoming publications. WFI2033−4723 is in the footprint of the Dark Energy Survey with observations in g, r, i, z and Y bands, so we did not target WFI2033−4723 in these bands. We observed only B1608+656 with *Spitzer* since the other four lenses have archival *Spitzer*/IRAC observations (PI: C. S. Kochanek).

6.5-m Magellan telescopes. In H0LiCOW Paper II, we merge the spectroscopic catalogue from the multiple spectroscopic campaigns on HE 0435−1223.

4.4 Lens galaxy spectroscopy for lens velocity dispersion

For B1608+656 and RXJ1131−1231, we have obtained long-slit spectra of the lens systems with the Low-Resolution Imaging Spectrometer (LRIS; Oke et al. 1995) at the Keck Observatory for measuring the lens stellar velocity dispersion (Suyu et al. 2010, 2013). For HE 0435−1223, we observe the lens system with LRIS in multiobject mode to obtain spectra of the foreground lens galaxy for lens velocity dispersion measurement (see H0LiCOW Paper IV) and also of nearby galaxies (see H0LiCOW Paper II). Both WFI2033−4723 and HE 1104−1805 have bright AGNs relative to the lens galaxy, making the lens velocity dispersion measurement challenging. We have new observations of WFI2033−4723 with MUSE (Bacon et al. 2012) at the VLT, which we expect will allow us to reduce the uncertainty on the current lens velocity dispersion by a factor of 2, to ∼5–7 per cent precision. The velocity dispersion is a key ingredient to break the MSD/lensing degeneracies (e.g. Suyu et al. 2014). For HE 1104−1805, we obtained one-sixth of our proposed observations with XSHOOTER on the VLT in priority B, which is not sufficient to measure the velocity dispersion. We have time on OSIRIS (OH-Suppressing Infra-Red Imaging Spectrograph; Larkin et al. 2006) on Keck to observe HE 1104−1805, RXJ1131−1231and HE 0435−1223 with adaptive optics. Because OSIRIS is an integral field spectrograph, these observations have the goal of obtaining two-dimensional kinematic data of the foreground lens, which will then be used to further constrain the lens mass models. We summarize the spectroscopic observations for lens velocity dispersion measurement in Table 4.

Table 4. Spectroscopy of foreground lens as part of the H0LiCOW program.

Lens	Facility/instrument	Proposal PI
B1608+656	W. M. Keck/LRIS	C. D. Fassnacht
RXJ1131−1231	W. M. Keck/LRIS	C. D. Fassnacht
	W. M. Keck/OSIRIS	T. Treu
HE 0435−1223	W. M. Keck/LRIS	C. D. Fassnacht
	W. M. Keck/OSIRIS	T. Treu
WFI2033−4723	VLT/MUSE	D. Sluse
HE 1104−1805	VLT/X-shooter	C. Spiniello
	W. M. Keck/OSIRIS	T. Treu

Notes. OSIRIS is the OH-Suppressing Infra-Red Imaging Spectrograph (Larkin et al. 2006). Details of the LRIS observations for B1608+656 are in Suyu et al. (2010), for RXJ1131−1231 are in Suyu et al. (2013), and for HE 0435−1223 are in H0LiCOW Paper IV; other observations are in forthcoming publications. Only one-sixth of the HE 1104−1805 observations with X-shooter (Vernet et al. 2011) were obtained, which were insufficient for measuring the lens velocity dispersion. The observations with OSIRIS are pending.

5 COSMOGRAPHY AND ASTROPHYSICS WITH HE 0435−1223: KEY COMPONENTS

We summarize the key ingredients and analysis of HE 0435−1223 that are described in upcoming publications of the H0LiCOW project (H0LiCOW Papers II–V). The titles of the papers begin with ‘H0LiCOW’, followed by the specific titles written below.

II. Spectroscopic survey and galaxy-group identification of the strong gravitational lens systems HE 0435−1223 (H0LiCOW Paper II). From our spectroscopic campaign of the lens environment, we present the measured spectroscopic redshifts, focusing in particular on the massive and nearby objects to the strong-lens system that are necessary ingredients for lens mass modelling and distance measurement. By combining with the spectroscopic catalogue of independent efforts (Momcheva et al. 2015), we identify potential galaxy groups towards HE 0435−1223 in order to control the systematic effect due to the galaxies along the line of sight. We use the flexion shift[7] introduced by McCully et al. (2016) to determine which mass structures (galaxies/groups) need to be incorporated explicitly in the lens mass model and which could be well approximated by an external shear/convergence field. The flexion-shift analysis presented in H0LiCOW Paper II shows that the most significant line-of-sight perturber is the galaxy G1 that is closest to the lens system, which justifies our inclusion of this particular galaxy in all of our strong-lensing models in H0LiCOW Paper IV. Furthermore, the next four nearest perturbers from the lens system may also produce higher order perturbations on the lens potential, and we account for the effects of these four additional galaxies in one of our systematic tests in H0LiCOW Paper IV.

III. Quantifying the effect of mass along the line of sight to the gravitational lens HE 0435−1223 through weighted galaxy counts (H0LiCOW Paper III). Using the wide-field photometry and spectroscopy in Section 4.3, we compute photometric redshifts and stellar masses for objects in the field up to 120 arcsec from the strong lens, and with $i < 24$ mag. We thoroughly test the weighted galaxy number counts technique of Greene et al. (2013), and apply it to HE 0435−1223 with the CFHTLenS survey (Heymans et al. 2012) as the control field. By comparing the weighted counts to simulated lines of sight that are ray traced through numerical simulations (Hilbert et al. 2007, 2009), we infer the distribution for the external convergence κ_{ext} that excludes the strong-lens redshift plane.

IV. Lens mass model of HE 0435−1223 and blind measurement of its time-delay distance for cosmology (H0LiCOW Paper IV). Using the time delays from H0LiCOW Paper V and our *HST*/WFC3-IR imaging (*F*160*W*) and archival *HST*/ACS observations (*F*555*W* and *F*814*W*), we model the lens mass distribution including explicitly the nearest, in projection from HE 0435−1223, one (G1) or five (G1 plus the next four nearest/brightest) perturbers, with spectroscopic redshifts from H0LiCOW Paper II. We then incorporate the velocity dispersion of the lens galaxy, and the external convergence distribution from H0LiCOW Paper III to infer an effective time-delay distance, which is blinded during the analysis stage. We unblind only after the completion of the analysis, and publish these results without modifications.

V. New COSMOGRAIL time delays of HE 0435−1223: H_0 to 3.8 per cent from strong lensing in flat-ΛCDM (H0LiCOW Paper V). We present the 13-yr monitoring of HE 0435−1223 and measure the time delays between the image pairs. Using the resulting effective time-delay distance of HE 0435−1223 from the blind analysis in H0LiCOW Paper IV, we create a Time Delay Strong Lensing (TDSL) probe with HE 0435−1223, RXJ1131−1231 and B1608+656 (we note that the analysis of RXJ1131−1231 was also blinded in Suyu et al. (2013), whereas the analysis of B1608+656 was not as it was the first lens to be analysed using our modelling techniques). We infer cosmological constraints from TDSL alone, and combine it with other cosmological probes to constrain various cosmological models.

In addition to the above, there are more forthcoming publications. The study of the AGN host galaxy properties based on simulations are described in H0LiCOW Paper VI (Ding et al. 2017). The newly developed multilens plane modelling, based on the multilens plane equations (Schneider et al. 1992; Blandford & Kochanek 2004), and PSF reconstruction will be detailed by Suyu et al. (in preparation). The weak-lensing analysis of the field of HE 0435−1223 will be presented by Tihhonova et al. (in preparation). Following these publications, there will be the next studies and analysis of the remaining sample (WFI2033−4723 and HE 1104−1805).

6 H0LICOW COSMOGRAPHIC FORECAST

We make predictions of the cosmographic constraints based on our sample of H0LiCOW lenses. We use the time-delay distance measurements for B1608+656 (equation 35 of Suyu et al. 2010), RXJ1131−1231 (equation 5 of Suyu et al. 2014) and HE 0435−1223 (equation 17 of H0LiCOW Paper IV). For the forecasted time-delay distance measurements of the other two lenses, we adopt an uncertainty with contributions from the time delays, mass modelling and external convergence added in quadrature. Specifically, we estimate time-delay uncertainties of 4 per cent and 2 per cent, modelling uncertainties of 4 per cent and 8 per cent, external convergence uncertainties of 4 per cent and 4 per cent, yielding a total uncertainty of 7 per cent and 9 per cent for WFI2033−4723 and HE 1104−1805, respectively. Furthermore, we assume that the angular diameter distance to each lens can be measured with an uncertainty of 15 per cent using our current data sets (Jee et al. 2015). More precise measurements of D_{d} (∼5–10 per cent uncertainty) would require additional kinematic data of the lens galaxy beyond what we currently have, particularly spatially resolved kinematics maps. For the forecasted $D_{\Delta t}$ and D_{d} constraints, we adopt a fiducial cosmological model with $H_0 = 72\,\text{km}\,\text{s}^{-1}\,\text{Mpc}^{-1}$, $\Omega_{\text{m}} = 1 - \Omega_{\text{DE}} = 0.32$, and $w = -1$ to predict the distances with their estimated uncertainties mentioned above, although we note that this assumption affects little the fractional uncertainty, which is nearly scale-free.

We show in Fig. 3, the cosmographic constraints of our sample of lenses with uniform priors on the cosmological parameters (left-column panels), in combination with *WMAP* 9-yr results (Hinshaw et al. 2013, middle-left-column panels), and in combination with *Planck* 2015 results (Planck Collaboration XIII 2016, middle-right-column panels)[8] for three different background cosmologies: (1) open ΛCDM with variable spatial curvature Ω_{k} (top row), (2) spatially flat *w*CDM with w as the time-independent dark energy equation of state (middle row) and (3) flat ΛCDM with varying effective number of relativistic species N_{eff} (bottom row). In the right-column panels, we show the one-dimensional marginalized constraints of H_0 of our sample of lenses alone or in combination with the CMB data sets (i.e. marginalized H_0 distributions of the panels to the left), as indicated in the legend. We list in Table 5 the prior ranges for the uniform background cosmologies. The *WMAP* 9-yr and *Planck*

---

[7] The flexion shift corresponds to the shift in the image positions due to the flexion (third-order derivatives of the lens potential) of a line-of-sight perturber. McCully et al. (2016) find through their study of simulated lens fields that perturbers with flexion shifts larger than $\sim 10^{-4}$ arcsec should be incorporated explicitly in the multiplane lens mass model. The threshold of $\sim 10^{-4}$ arcsec is conservative and is based on tests that only used image positions as constraints. It may be that using the spatially extended images for modelling would push that threshold even lower.

[8] We use the *Planck* chains designated by 'plikHM_TT_lowTEB' that uses the baseline high-L *Planck* power spectra and low-L temperature and LFI polarization.

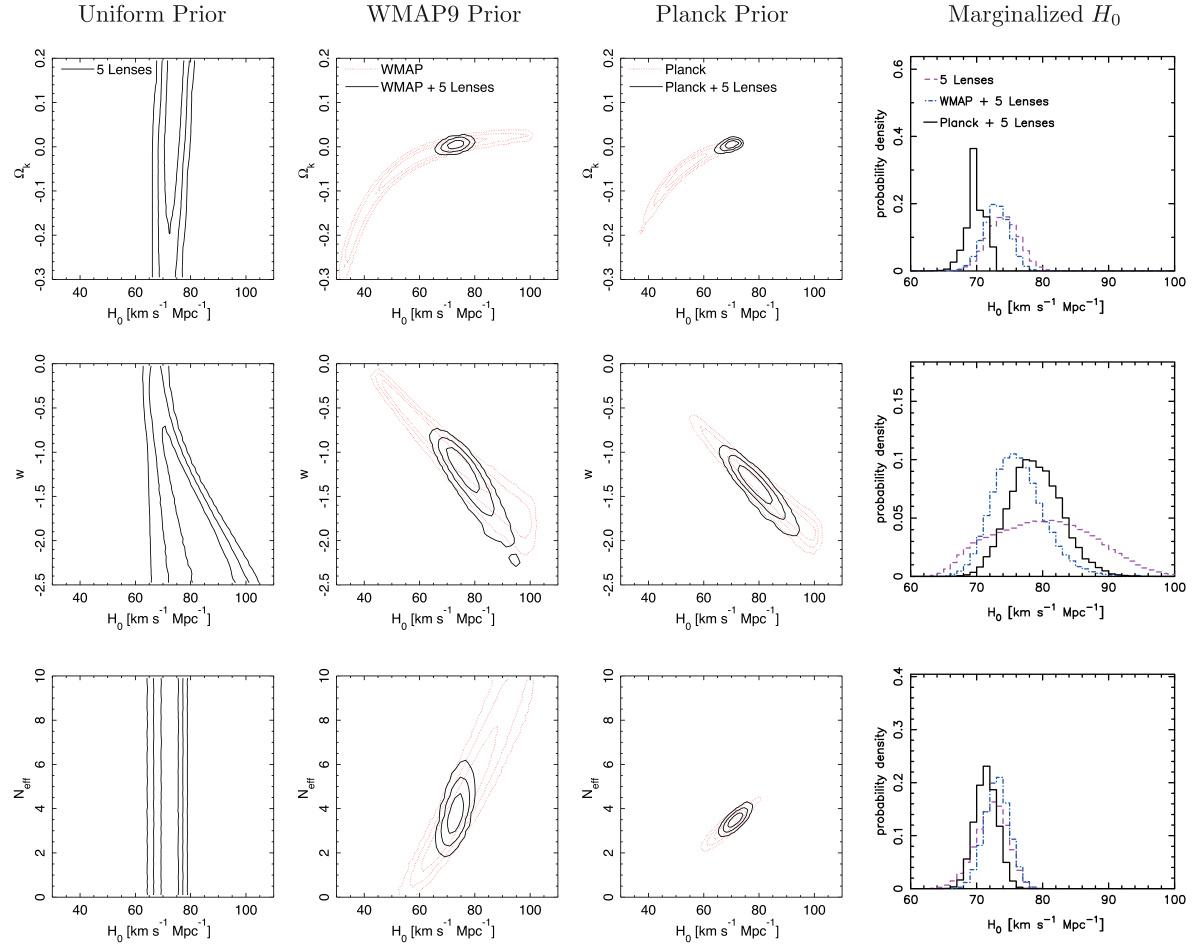


Figure 3. Forecasted cosmographic constraints from the H0LiCOW lens sample through measurements of $D_{\Delta t}$ and D_d. Columns from left to right are, respectively, the constraints from the H0LiCOW lenses alone (with uniform prior on cosmological parameters), lenses in combination with *WMAP* 9-yr results, lenses in combination with *Planck* 2015 results, and marginalized constraints on H_0 from the previous three columns. The H0LiCOW lenses primarily constrain H_0, which in turn break CMB parameter degeneracies to elucidate the spatial curvature of universe (Ω_k, top row), dark energy equation of state (w, middle row) and effective number of relativistic species ($N_{\rm eff}$, bottom row). H0LiCOW lenses provide an independent, complementary and competitive probe of cosmology.

Table 5. Prior for 'uniform' cosmological models.

Cosmology	Prior ranges
Open ΛCDM	$H_0 \in [0, 120]$ km s^{-1} Mpc^{-1} $\Omega_m \in [0, 0.5]$ $\Omega_\Lambda \in [0.5, 1]$ $\Omega_k = 1 - \Omega_m - \Omega_\Lambda$
Flat *w*CDM	$H_0 \in [0, 120]$ km s^{-1} Mpc^{-1} $\Omega_m \in [0, 1]$ $\Omega_{DE} = 1 - \Omega_m$ $w \in [-2.5, 0]$
Flat $N_{\rm eff}$ΛCDM	$H_0 \in [0, 120]$ km s^{-1} Mpc^{-1} $\Omega_m \in [0, 1]$ $\Omega_\Lambda = 1 - \Omega_m$ $N_{\rm eff} \in [0, 10]$

chains have a prior with $H_0 < 100\,{\rm km\,s^{-1}\,Mpc^{-1}}$ imposed. The cosmographic constraints of our lenses shown in Fig. 3 (from the forecasted measurements of $D_{\Delta t}$ and D_d) mostly stem from the $D_{\Delta t}$ measurements as a results of the substantially smaller uncertainties of $D_{\Delta t}$ than that of D_d. In fact, the cosmographic constraints from $D_{\Delta t}$ alone would increase the H_0 1σ uncertainties shown in Fig. 3 by at most $0.8\,{\rm km\,s^{-1}\,Mpc^{-1}}$ (depending on the background cosmology). The additional cosmographic information from D_d would become more significant when the D_d uncertainties are reduced to ∼5–10 per cent (Jee et al. 2016).

As seen in the left column, the time-delay lenses primarily constrain H_0, and depend weakly (if at all) on other parameters. None the less, the time-delay distances $D_{\Delta t}$ and the lenses' angular diameter distances D_d provide some information on w, as the constraint contours are tilted rather than being vertical. With more lenses or smaller uncertainties on D_d measurements, the constraints on cosmology become more prominent (Jee et al. 2016). However, the

H0LiCOW lenses provide strong cosmographic constraints when combined with the CMB measurements since they help to break parameter degeneracies in the CMB. Thus, we should be able to place substantially better constraints on, for example, the spatial curvature, w and N_{eff} (middle two columns), compared to constraints from CMB alone. In particular, we expect better than 3.5 per cent precision on H_0 for the two cosmologies with $w = -1$ (open ΛCDM and flat N_{eff}ΛCDM)[9]; when w is allowed to vary, this constraint weakens to ∼11 per cent without CMB priors and ∼5 per cent with CMB priors in the wCDM cosmology, as visible in the rightmost panel in the middle row. By combining our five H0LiCOW lenses with *Planck*, we expect to achieve the following precisions: Ω_{k} to 0.004 in open ΛCDM, w to 0.14 in flat wCDM, and N_{eff} to 0.2 in flat N_{eff}ΛCDM (all 1σ uncertainties). These precisions are a factor of ∼15, ∼2, and ∼1.5, respectively, tighter than *Planck* on its own. Our H0LiCOW sample provides not only an independent check of systematics, but also a great complement to other cosmological probes for pinning down cosmological parameters.

7 SUMMARY AND OUTLOOK

We present the H0LiCOW program that aims to measure H_0 to <3.5 per cent in precision and accuracy (in most background cosmological models) with a sample of five time-delay lenses, completely independent of the cosmic distance ladder and other direct measurements of H_0. Our cosmographic information comes from measuring the distances to the lens systems, specifically $D_{\Delta t}$ and D_{d}.

To achieve our goal, we have obtained almost all the key ingredients for our lens sample[10]: (1) the time delays from the COSMOGRAIL and Very Large Array monitoring, (2) high-resolution *HST* imaging for modelling the lens mass distributions, (3) wide-field imaging and spectroscopy to quantify the effects of the lens environment, and (4) lens velocity dispersion measurements to augment our lensing mass models. Our new *HST* observations reveal Einstein rings in the lens systems that allow us to perform precision lens mass modelling.

The results of our recent blind analysis of HE 0435−1223 will appear in the companion H0LiCOW publications. H0LiCOW Paper II (Sluse et al. 2017) presents the spectroscopic campaign on the HE 0435−1223 field and identifies galaxy groups in the light cone containing the lens. H0LiCOW Paper III (Rusu et al. 2017) combines the spectroscopy, the wide-field imaging data, and the Millennium Simulation to derive the external convergence of the line-of-sight mass distributions. H0LiCOW Paper IV (Wong et al. 2017) models the lens mass distribution using the *HST* data, the time delays and the lens velocity dispersion to infer the time-delay distance, that is blinded throughout the analysis. H0LiCOW Paper V (Bonvin et al. 2017) presents the COSMOGRAIL monitoring of HE 0435−1223 and investigates the cosmological implications based on the three lenses (B1608+656, RXJ1131−1231and HE 0435−1223) that we have so far analysed.

With our sample of five lenses, we expect to measure H_0 to <3.5 per cent in precision and accuracy for the non-flat ΛCDM cosmology or flat N_{eff}ΛCDM cosmology, with $w = -1$. When w is allowed to vary, the constraint on H_0 degrades to ∼11 per cent with time-delay data only, and to ∼5 per cent when augmented with CMB data. Our independent strong-lensing distances significantly improve cosmological constraints from the *Planck* data: the precisions on Ω_{k}, w, and N_{eff} improve by a factor of ∼15, ∼2, and 1.5, respectively, when we combine our lenses with *Planck*. Time-delay lenses are therefore highly complementary to other cosmological probes.

Our data set provides an excellent opportunity to study, in addition to cosmography, galaxy formation, and evolution. For example, we can study the distribution of dark matter in the lens galaxies by combining lensing and kinematics data, and also infer the stellar mass of the lens galaxies (e.g. Treu & Koopmans 2004; Barnabè et al. 2011; Sonnenfeld et al. 2012, 2015; Suyu et al. 2012b). By separately determining the stellar mass based on either (1) stellar population synthesis using multiband photometry (e.g. Auger et al. 2009; Treu et al. 2010; Oguri, Rusu & Falco 2014), or (2) identification/characterization of spectral features (e.g. van Dokkum & Conroy 2010; Conroy & van Dokkum 2012; Spiniello et al. 2012, 2014, 2015; Barnabè et al. 2013), and comparing this stellar mass to that obtained from lensing and dynamics, we can study properties of the stellar population and infer the stellar IMF slope (e.g. Grillo et al. 2009; Auger et al. 2010; Treu et al. 2010; Spiniello et al. 2011, 2015; Barnabè et al. 2013). There are about a dozen early-type lens galaxies that have been studied in detail for constraining the stellar IMF slope individually (e.g. Sonnenfeld et al. 2012; Barnabè et al. 2013; Spiniello et al. 2015; Newman et al. 2016), and these galaxies are all at redshifts below 0.35. Four of our H0LiCOW lens galaxies are at redshifts between 0.45 and 0.73, which would allow us to explore the stellar IMF with comparable precisions per lens galaxy as previous studies, but at substantially higher redshifts. Given the current tension in the IMF measurement between nearby ($z_{\mathrm{d}} < 0.06$) lens galaxies and $z_{\mathrm{d}} \sim 0.2$–0.3 lens galaxies (e.g. Smith & Lucey 2013; Newman et al. 2016), our H0LiCOW lenses would help assess whether the tensions are just limited to those particular objects or if they reflect a more general problem in our understanding of stellar populations. In addition, our lenses are natural telescopes that magnify the background sources, allowing us to study the host galaxies of the AGNs in detail and probe the origin of the co-evolution between supermassive black holes and their host galaxies (Peng et al. 2006; Rusu et al. 2016; Ding et al. 2017).

Our H0LiCOW program aims to establish gravitational lens time delays as an independent and competitive probe of cosmology, and paves the way for determining H_0 to 1 per cent in the future. Given the hundreds, if not thousands, of time-delay lens systems that are expected to be discovered in ongoing and future surveys such as the Sloan Digital Sky Survey (e.g. Oguri et al. 2006; Inada et al. 2012; More et al. 2016a), the Dark Energy Survey (e.g. Agnello et al. 2015), the Hyper Suprime-Cam Survey (e.g. Chan et al. 2016), the Kilo-Degree Survey (e.g. Napolitano et al. 2015), Euclid and the Large Synoptic Survey Telescope (Oguri & Marshall 2010), and continuous advances in high-resolution imaging and spectroscopy in the current and next generation of telescopes for observational follow-up (Linder 2015; Meng et al. 2015), the H0LiCOW program will provide the basis for extracting cosmological information from the wealth of strong-lensing data sets. In particular, we expect the combination of facilities at different wavelengths such as the *HST* in the optical/near-IR, *James Webb Space Telescope* in the IR, large and extremely large telescopes with adaptive optics, the Atacama Large Millimeter/submillimeter Array in the submillimetre waveband, and the Square Kilometer Array in the radio, will be of great synergistic value for studying these fruitful lenses.

---

[9] Relative to $H_0 = 72\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

[10] With spectroscopic observations of HE 1104−1805 pending for lens velocity dispersion measurement.

ACKNOWLEDGEMENTS

We thank M. Barnabè, J. Chan, Y. Hezaveh, E. Komatsu, E. Linder, J. McKean, D. Paraficz, P. Schneider and S. Vegetti for helpful discussions, and the anonymous referee for detailed comments that improved the presentation of this work. H0LiCOW and COSMOGRAIL are made possible thanks to the continuous work of all observers and technical staff obtaining the monitoring observations, in particular at the Swiss Euler telescope at La Silla Observatory. Euler is supported by the Swiss National Science Foundation. SHS is supported by the Max Planck Society through the Max Planck Research Group. This work is supported in part by the Ministry of Science and Technology in Taiwan via grant MOST-103-2112-M-001-003-MY3. VB, FC and GM acknowledge the support of the Swiss National Science Foundation (SNSF). CDF and CER were funded through the NSF grant AST-1312329, 'Collaborative Research: Accurate cosmology with strong gravitational lens time delays', and CDF and NR were funded by the *HST* grant GO-12889. DS acknowledges funding support from a *Back to Belgium* grant from the Belgian Federal Science Policy (BELSPO). TT thanks the Packard Foundation for generous support through a Packard Research Fellowship, the NSF for funding through NSF grant AST-1450141, 'Collaborative Research: Accurate cosmology with strong gravitational lens time delays'. KCW is supported by an EACOA Fellowship awarded by the East Asia Core Observatories Association, which consists of the Academia Sinica Institute of Astronomy and Astrophysics, the National Astronomical Observatory of Japan, the National Astronomical Observatories of the Chinese Academy of Sciences, and the Korea Astronomy and Space Science Institute. XD is supported by the China Scholarship Council. SH acknowledges support by the DFG cluster of excellence 'Origin and Structure of the Universe' (www.universe-cluster.de). PJM acknowledges support from the U.S. Department of Energy under contract number DE-AC02-76SF00515. MT acknowledges support by a fellowship of the Alexander von Humboldt Foundation and the DFG grant Hi 1495/2-1. LVEK is supported in part through an NWO-VICI career grant (project number 639.043.308). Based on observations made with the NASA/ESA *Hubble Space Telescope*, obtained at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. These observations are associated with program nos. 12889, 10158, 7422 and 9744. Support for program no. 12889 was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. Some of the data presented herein were obtained at the W.M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California and the National Aeronautics and Space Administration. The Observatory was made possible by the generous financial support of the W.M. Keck Foundation. The authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Mauna Kea has always had within the indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this mountain. Access to the CFHT was made possible by the Institute of Astronomy and Astrophysics, Academia Sinica, National Tsing Hua University, and National Science Council, Taiwan. Based in part on data collected at Subaru Telescope, which is operated by the National Astronomical Observatory of Japan. Based on observations obtained at the Gemini Observatory, which is operated by the Association of Universities for Research in Astronomy, Inc., under a cooperative agreement with the NSF on behalf of the Gemini partnership: the National Science Foundation (United States), the National Research Council (Canada), CONICYT (Chile), Ministerio de Ciencia, Tecnología e Innovación Productiva (Argentina), and Ministério da Ciência, Tecnologia e Inovação (Brazil). These observations are associated with programs GN-2012B-Q-11, GN-2013A-Q-72, GS-2013A-Q-2 and GS-2013B-Q-28. Based on observations made with ESO Telescopes at the La Silla Paranal Observatory under programme ID 090.A-0531(A), P91.B-0346(B), 091.A-0642(A), 092.A-0515(A), 092.A-0515(B), 60.A-9306(A) and 097.A-0454(A). This work is based in part on observations and archival data obtained with the *Spitzer Space Telescope*, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA. This project used data obtained with the Dark Energy Camera (DECam), which was constructed by the Dark Energy Survey (DES) collaboration. Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovacão, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey. The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Enérgeticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, the National Optical Astronomy Observatory, the University of Nottingham, the Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex and Texas A&M University.

REFERENCES

Agnello A. et al., 2015, MNRAS, 454, 1260

Anderson L. et al., 2014, MNRAS, 441, 24

Appenzeller I. et al., 1998, The Messenger, 94, 1

Aubourg É. et al., 2015, Phys. Rev. D, 92, 123516

Auger M. W., Treu T., Bolton A. S., Gavazzi R., Koopmans L. V. E., Marshall P. J., Bundy K., Moustakas L. A., 2009, ApJ, 705, 1099

Auger M. W., Treu T., Bolton A. S., Gavazzi R., Koopmans L. V. E., Marshall P. J., Moustakas L. A., Burles S., 2010, ApJ, 724, 511

Bacon R. et al., 2012, The Messenger, 147, 4

Barnabè M., Czoske O., Koopmans L. V. E., Treu T., Bolton A. S., Gavazzi R., 2009, MNRAS, 399, 21

Barnabè M., Czoske O., Koopmans L. V. E., Treu T., Bolton A. S., 2011, MNRAS, 415, 2215

Barnabè M., Spiniello C., Koopmans L. V. E., Trager S. C., Czoske O., Treu T., 2013, MNRAS, 436, 253

Beaton R. L. et al., 2016, ApJ, 832, 210

Betoule M. et al., 2014, A&A, 568, A22

Birrer S., Amara A., Refregier A., 2015, ApJ, 813, 102

Birrer S., Amara A., Refregier A., 2016, JCAP, 08, 020

Blake C. et al., 2011, MNRAS, 415, 2892

Blandford R. D., Kochanek C. S., 2004, in Bahcall J. et al., eds, Gravitational Lenses. World Scientific Press, Singapore, p. 103

Blandford R., Narayan R., 1986, ApJ, 310, 568

Bonvin V., Tewes M., Courbin F., Kuntzer T., Sluse D., Meylan G., 2016, A&A, 585, A88

Bonvin V. et al., 2017, MNRAS 465, 4914 (H0LiCOW Paper V)

Boulade O. et al., 2003, in Iye M., Moorwood A. F. M., eds, Proc. SPIEVol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes. SPIE, Bellingham, p. 72

Browne I. W. A. et al., 2003, MNRAS, 341, 13

Burud I. et al., 2002, A&A, 383, 71

Cantale N., Courbin F., Tewes M., Jablonka. P., Meylan G., 2016, A&A, 589, A81

Casali M. et al., 2006, in McLean I. S., Iye M., eds, Proc. SPIEConf. Ser. Vol. 6269, Ground-based and Airborne Instrumentation for Astronomy. SPIE, Bellingham, p. 62690W

Chan J. H. H. et al., 2016, ApJ, 832, 135

Chen G. C. F. et al., 2016, MNRAS, 462, 3457

Claeskens J.-F., Sluse D., Riaud P., Surdej J., 2006, A&A, 451, 865

Collett T. E., Cunnington S. D., 2016, MNRAS, 462, 3255

Collett T. E. et al., 2013, MNRAS, 432, 679

Conley A. et al., 2006, ApJ, 644, 1

Conley A. et al., 2011, ApJS, 192, 1

Conroy C., van Dokkum P. G., 2012, ApJ, 760, 71

Courbin F., Lidman C., Magain P., 1998, A&A, 330, 57

Courbin F., Eigenbrod A., Vuissoz C., Meylan G., Magain P., 2005, in Mellier Y., Meylan G., eds, Proc. IAU Symp. 225, Gravitational Lensing Impact on Cosmology. Cambridge Univ. Press, Cambridge, p. 297

Courbin F. et al., 2011, A&A, 536, A53

Dark Energy Survey Collaboration, Diehl T., 2012, Phys. Procedia, 37, 1332

Ding X. et al., 2017, MNRAS, 465, 4634

Dobler G., Fassnacht C. D., Treu T., Marshall P., Liao K., Hojjati A., Linder E., Rumbaugh N., 2015, ApJ, 799, 168

Dye S., Evans N. W., Belokurov V., Warren S. J., Hewett P., 2008, MNRAS, 388, 384

Eigenbrod A., Courbin F., Meylan G., Vuissoz C., Magain P., 2006, A&A, 451, 759

Eisenstein D. J. et al., 2005, ApJ, 633, 560

Falco E. E., Gorenstein M. V., Shapiro I. I., 1985, ApJ, 289, L1

Fassnacht C. D., Womble D. S., Neugebauer G., Browne I. W. A., Readhead A. C. S., Matthews K., Pearson T. J., 1996, ApJ, 460, L103

Fassnacht C. D., Pearson T. J., Readhead A. C. S., Browne I. W. A., Koopmans L. V. E., Myers S. T., Wilkinson P. N., 1999, ApJ, 527, 498

Fassnacht C. D., Xanthopoulos E., Koopmans L. V. E., Rusin D., 2002, ApJ, 581, 823

Fassnacht C. D., Gal R. R., Lubin L. M., McKean J. P., Squires G. K., Readhead A. C. S., 2006, ApJ, 642, 30

Fassnacht C. D., Koopmans L. V. E., Wong K. C., 2011, MNRAS, 410, 2167

Fazio G. G. et al., 2004, ApJS, 154, 10

Freedman W. L., Madore B. F., Scowcroft V., Burns C., Monson A., Persson S. E., Seibert M., Rigby J., 2012, ApJ, 758, 24

Futamase T., Hamana T., 1999, Prog. Theor. Phys., 102, 1037

Futamase T., Yoshida S., 2001, Prog. Theor. Phys., 105, 887

Gao F. et al., 2016, ApJ, 817, 128

Goobar A. et al., 2016, preprint (arXiv:1611.00014)

Greene Z. S. et al., 2013, ApJ, 768, 39

Grillo C., Lombardi M., Bertin G., 2008, A&A, 477, 397

Grillo C., Gobat R., Lombardi M., Rosati P., 2009, A&A, 501, 461

Grillo C. et al., 2016, ApJ, 822, 78

Harva M., Raychaudhury S., 2008, Neurocomputing, 72, 32

Heavens A., Jimenez R., Verde L., 2014, Phys. Rev. Lett., 113, 241302

Hewett P. C., Wild V., 2010, MNRAS, 405, 2302

Heymans C. et al., 2012, MNRAS, 427, 146

Hilbert S., White S. D. M., Hartlap J., Schneider P., 2007, MNRAS, 382, 121

Hilbert S., Hartlap J., White S. D. M., Schneider P., 2009, A&A, 499, 31

Hinshaw G. et al., 2013, ApJS, 208, 19

Hirv A., Olspert N., Pelt J., 2011, Balt. Astron., 20, 125

Hjorth J. et al., 2002, ApJ, 572, L11

Hodapp K. W. et al., 2003, PASP, 115, 1388

Hojjati A., Kim A. G., Linder E. V., 2013, Phys. Rev. D, 87, 123512

Holder G. P., Schechter P. L., 2003, ApJ, 589, 688

Hook I. M., Jørgensen I., Allington-Smith J. R., Davies R. L., Metcalfe N., Murowinski R. G., Crampton D., 2004, PASP, 116, 425

Hu W., 2005, in Wolff S. C., Lauer T. R., eds, ASP Conf. Ser. Vol. 339, Observing Dark Energy. Astron. Soc. Pac., San Francisco, p. 215

Ichikawa T. et al., 2006, in McLean I. S., Iye M., eds, Proc. SPIEConf. Ser. Vol. 6269, Ground-based and Airborne Instrumentation for Astronomy. SPIE, Bellingham, p. 626916

Inada N. et al., 2012, AJ, 143, 119

Jee I., Komatsu E., Suyu S. H., 2015, J. Cosmology Astropart. Phys., 11, 033

Jee I., Komatsu E., Suyu S. H., Huterer D., 2016, J. Cosmology Astropart. Phys., 2016, 031

Kawamata R., Oguri M., Ishigaki M., Shimasaku K., Ouchi M., 2016, ApJ, 819, 114

Kazin E. A. et al., 2014, MNRAS, 441, 3524

Kelly P. L. et al., 2015, Science, 347, 1123

Kelly P. L. et al., 2016, ApJ, 819, L8

Kissler-Patig M. et al., 2008, A&A, 491, 941

Kochanek C. S., 2002, ApJ, 578, 25

Kochanek C. S., Keeton C. R., McLeod B. A., 2001, ApJ, 547, 50

Kochanek C. S., Morgan N. D., Falco E. E., McLeod B. A., Winn J. N., Dembicky J., Ketzeback B., 2006, ApJ, 640, 47

Komatsu E. et al., 2011, ApJS, 192, 18

Koopmans L. V. E., 2005, MNRAS, 363, 1136

Koopmans L. V. E., Treu T., Fassnacht C. D., Blandford R. D., Surpi G., 2003, ApJ, 599, 70

Kuo C. Y. et al., 2015, ApJ, 800, 26

Larkin J. et al., 2006, in McLean I. S., Iye M., eds, Proc. SPIEConf. Ser. Vol. 6269, Ground-based and Airborne Instrumentation for Astronomy. SPIE, Bellingham, p. 62691A

Liao K. et al., 2015, ApJ, 800, 11

Lidman C., Courbin F., Kneib J.-P., Golse G., Castander F., Soucail G., 2000, A&A, 364, L62

Linder E. V., 2015, Phys. Rev. D, 91, 083511

McCully C., Keeton C. R., Wong K. C., Zabludoff A. I., 2014, MNRAS, 443, 3631

McCully C., Keeton C. R., Wong K. C., Zabludoff A. I., 2016, ApJ, 836, 141

Magain P., Courbin F., Sohy S., 1998, ApJ, 494, 472

Meng X.-L., Treu T., Agnello A., Auger M. W., Liao K., Marshall P. J., 2015, J. Cosmology Astropart. Phys., 9, 059

Miyazaki S. et al., 2002, PASJ, 54, 833

Momcheva I., Williams K., Keeton C., Zabludoff A., 2006, ApJ, 641, 169

Momcheva I. G., Williams K. A., Cool R. J., Keeton C. R., Zabludoff A. I., 2015, ApJS, 219, 29

More A. et al., 2016a, MNRAS, 456, 1595

More A., Suyu S. H., Oguri M., More S., Lee C.-H., 2016b, ApJ, 835, L25

Morgan N. D., Caldwell J. A. R., Schechter P. L., Dressler A., Egami E., Rix H.-W., 2004, AJ, 127, 2617

Morgan N. D., Kochanek C. S., Pevunova O., Schechter P. L., 2005, AJ, 129, 2531

Morgan C. W., Eyler M. E., Kochanek C. S., Morgan N. D., Falco E. E., Vuissoz C., Courbin F., Meylan G., 2008, ApJ, 676, 80

Myers S. T. et al., 1995, ApJ, 447, L5

Myers S. T. et al., 2003, MNRAS, 341, 1

Napolitano N. R. et al., 2015, The Universe of Digital Sky Surveys. Springer-Verlag, Berlin

Newman A. B., Smith R. J., Conroy C., Villaume A., van Dokkum P., 2016, ApJ, preprint (arXiv:1612.00065)
Ofek E. O., Maoz D., Rix H.-W., Kochanek C. S., Falco E. E., 2006, ApJ, 641, 70
Oguri M., Marshall P. J., 2010, MNRAS, 405, 2579
Oguri M. et al., 2006, AJ, 132, 999
Oguri M., Rusu C. E., Falco E. E., 2014, MNRAS, 439, 2494
Oke J. B. et al., 1995, PASP, 107, 375
Paraficz D., Hjorth J., 2009, A&A, 507, L49
Pelt J., Kayser R., Refsdal S., Schramm T., 1996, A&A, 305, 97
Peng C. Y., Impey C. D., Rix H.-W., Kochanek C. S., Keeton C. R., Falco E. E., Lehár J., McLeod B. A., 2006, ApJ, 649, 616
Percival W. J. et al., 2010, MNRAS, 401, 2148
Perlmutter S. et al., 1999, ApJ, 517, 565
Pirard J.-F. et al., 2004, in Moorwood A. F. M., Iye M., eds, Proc. SPIEConf. Ser. Vol. 5492, Ground-based Instrumentation for Astronomy. SPIE, Bellingham, p. 1763
Planck Collaboration XVI, 2014, A&A, 571, A16
Planck Collaboration XIII, 2016, A&A, 594, A13
Poindexter S., Morgan N., Kochanek C. S., Falco E. E., 2007, ApJ, 660, 146
Press W. H., Rybicki G. B., Hewitt J. N., 1992, ApJ, 385, 404
Quimby R. M. et al., 2014, Science, 344, 396
Refsdal S., 1964, MNRAS, 128, 307
Reid M. J., Braatz J. A., Condon J. J., Lo K. Y., Kuo C. Y., Impellizzeri C. M. V., Henkel C., 2013, ApJ, 767, 154
Remy M., Claeskens J.-F., Surdej J., Hjorth J., Refsdal S., Wucknitz O., Sørensen A. N., Grundahl F., 1998, New Astron., 3, 379
Riess A. G. et al., 1998, AJ, 116, 1009
Riess A. G. et al., 2016, ApJ, 826, 56
Rockosi C. et al., 2010, in McLean I. S., Ramsay S. K., Takami H., eds, Ground-based and Airborne Instrumentation for Astronomy III. SPIE, Bellingham, p. 77350R
Romanowsky A. J., Kochanek C. S., 1999, ApJ, 516, 18
Ross A. J., Samushia L., Howlett C., Percival W. J., Burden A., Manera M., 2015, MNRAS, 449, 835
Rumbaugh N., Fassnacht C. D., McKean J. P., Koopmans L. V. E., Auger M. W., Suyu S. H., 2015, MNRAS, 450, 1042
Rusu C. E. et al., 2016, MNRAS, 458, 2
Rusu C. E. et al., 2017, preprint (arXiv:1607.01047) (H0LiCOW Paper III)
Schechter P. L. et al., 1997, ApJ, 475, L85
Schneider P., Sluse D., 2013, A&A, 559, A37
Schneider P., Sluse D., 2014, A&A, 564, A103
Schneider P., Ehlers J., Falco E. E., 1992, Gravitational Lenses, XIV. Springer-Verlag, Berlin
Sheinis A. I., Bolte M., Epps H. W., Kibrick R. I., Miller J. S., Radovan M. V., Bigelow B. C., Sutin B. M., 2002, PASP, 114, 851
Sluse D., Tewes M., 2014, A&A, 571, A60
Sluse D. et al., 2003, A&A, 406, L43
Sluse D., Claeskens J.-F., Hutsemékers D., Surdej J., 2007, A&A, 468, 885
Sluse D., Hutsemékers D., Courbin F., Meylan G., Wambsganss J., 2012, A&A, 544, A62
Sluse D. et al., 2017, MNRAS, preprint (arXiv:1607.00382) (H0LiCOW Paper II)
Smette A., Robertson J. G., Shaver P. A., Reimers D., Wisotzki L., Koehler T., 1995, A&AS, 113, 199
Smith R. J., Lucey J. R., 2013, MNRAS, 434, 1964
Sonnenfeld A., Treu T., Gavazzi R., Marshall P. J., Auger M. W., Suyu S. H., Koopmans L. V. E., Bolton A. S., 2012, ApJ, 752, 163
Sonnenfeld A., Treu T., Marshall P. J., Suyu S. H., Gavazzi R., Auger M. W., Nipoti C., 2015, ApJ, 800, 94
Spiniello C., Koopmans L. V. E., Trager S. C., Czoske O., Treu T., 2011, MNRAS, 417, 3000
Spiniello C., Trager S. C., Koopmans L. V. E., Chen Y. P., 2012, ApJ, 753, L32
Spiniello C., Trager S., Koopmans L. V. E., Conroy C., 2014, MNRAS, 438, 1483
Spiniello C., Barnabè M., Koopmans L. V. E., Trager S. C., 2015, MNRAS, 452, L21
Surpi G., Blandford R. D., 2003, ApJ, 584, 100
Suyu S. H., 2012, MNRAS, 426, 868
Suyu S. H., Halkola A., 2010, A&A, 524, A94
Suyu S. H., Marshall P. J., Blandford R. D., Fassnacht C. D., Koopmans L. V. E., McKean J. P., Treu T., 2009, ApJ, 691, 277
Suyu S. H., Marshall P. J., Auger M. W., Hilbert S., Blandford R. D., Koopmans L. V. E., Fassnacht C. D., Treu T., 2010, ApJ, 711, 201
Suyu S. H. et al., 2012a, preprint (arXiv:1202.4459)
Suyu S. H. et al., 2012b, ApJ, 750, 10
Suyu S. H. et al., 2013, ApJ, 766, 70
Suyu S. H. et al., 2014, ApJ, 788, L35
Suzuki R. et al., 2008, PASJ, 60, 1347
Suzuki N. et al., 2012, ApJ, 746, 85
Tewes M., Courbin F., Meylan G., 2013a, A&A, 553, A120
Tewes M. et al., 2013b, A&A, 556, A22
Treu T., Koopmans L. V. E., 2002, MNRAS, 337, L6
Treu T., Koopmans L. V. E., 2004, ApJ, 611, 739
Treu T., Marshall P. J., 2016, A&AR, 24, 11
Treu T., Gavazzi R., Gorecki A., Marshall P. J., Koopmans L. V. E., Bolton A. S., Moustakas L. A., Burles S., 2009, ApJ, 690, 670
Treu T., Auger M. W., Koopmans L. V. E., Gavazzi R., Marshall P. J., Bolton A. S., 2010, ApJ, 709, 1195
Treu T. et al., 2016, ApJ, 817, 60
Unruh S., Schneider P., Sluse D., 2016, A&A, preprint (arXiv:1606.04321)
van Dokkum P. G., Conroy C., 2010, Nature, 468, 940
Vanderriest C., Schneider J., Herpe G., Chevreton M., Moles M., Wlerick G., 1989, A&A, 215, 1
Vegetti S., Koopmans L. V. E., 2009, MNRAS, 392, 945
Vernet J. et al., 2011, A&A, 536, A105
von der Linden A. et al., 2014, MNRAS, 439, 2
Vuissoz C. et al., 2007, A&A, 464, 845
Vuissoz C. et al., 2008, A&A, 488, 481
Walsh D., Carswell R. F., Weymann R. J., 1979, Nature, 279, 381
Warren S. J., Dye S., 2003, ApJ, 590, 673
Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, Phys. Rep., 530, 87
Williams K. A., Momcheva I., Keeton C. R., Zabludoff A. I., Lehár J., 2006, ApJ, 646, 85
Wisotzki L., Koehler T., Kayser R., Reimers D., 1993, A&A, 278, L15
Wisotzki L., Christlieb N., Bade N., Beckmann V., Köhler T., Vanelle C., Reimers D., 2000, A&A, 358, 77
Wisotzki L., Schechter P. L., Bradt H. V., Heinmüller J., Reimers D., 2002, A&A, 395, 17
Wong K. C. et al., 2017, MNRAS, 465, 4895 (H0LiCOW Paper IV)
Wucknitz O., 2002, MNRAS, 332, 951
Wucknitz O., Biggs A. D., Browne I. W. A., 2004, MNRAS, 349, 14
Xu D., Sluse D., Schneider P., Springel V., Vogelsberger M., Nelson D., Hernquist L., 2016, MNRAS, 456, 739

[1] *Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, D-85748 Garching, Germany*
[2] *Institute of Astronomy and Astrophysics, Academia Sinica, PO Box 23-141, Taipei 10617, Taiwan*
[3] *Physik-Department, Technische Universität München, James-Franck-Straße 1, D-85748 Garching, Germany*
[4] *Laboratoire d'Astrophysique, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, CH-1290 Versoix, Switzerland*
[5] *Department of Physics, University of California, Davis, CA 95616, USA*
[6] *STAR Institute, Quartier Agora – Allée du six Août, 19c B-4000 Liège, Belgium*
[7] *Department of Physics and Astronomy, University of California, Los Angeles, CA 90095, USA*
[8] *National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan*
[9] *Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*

[10]*Department of Astronomy, Beijing Normal University, Beijing 100875, China*
[11]*Exzellenzcluster Universe, Boltzmannstr. 2, D-85748 Garching, Germany*
[12]*Ludwig-Maximilians-Universität, Universitäts-Sternwarte, Scheinerstr. 1, D-81679 München, Germany*
[13]*Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 452 Lomita Mall, Stanford, CA 94035, USA*
[14]*Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan*
[15]*Physics Department, University of California, Santa Barbara, CA 93106, USA*
[16]*Argelander-Institut für Astronomie, Auf dem Hügel 71, D-53121 Bonn, Germany*
[17]*ESO – European Southern Observatory, D-85748 Garching bei München, Germany*
[18]*Institute of Cosmology and Gravitation, University of Portsmouth, Burnaby Rd, Portsmouth PO1 3FX, UK*
[19]*Kapteyn Astronomical Institute, University of Groningen, PO Box 800, NL-9700 AV Groningen, the Netherlands*

This paper has been typeset from a TEX/LATEX file prepared by the author.

A Additional publications

Mon. Not. R. Astron. Soc. **000**, 1–14 (0000) Printed 24 June 2017 (MN LaTeX style file v2.2)

A. Agnello[1⋆], H. Lin[2], L. Buckley-Geer[2], T. Treu[3†], V. Bonvin[4], F. Courbin[4],
C. Lemon[5], T. Morishita[3], A. Amara[6], M.W. Auger[5], S. Birrer[6,3], J. Chan[7,8],
T. Collett[9], A. More[10], C.D. Fassnacht[11], J. Frieman[2], P.J. Marshall[12], R.G. McMahon[5,13],
G. Meylan[4], S.H. Suyu[8], F.Castander[14], D. Finley[2], A.Howell[15],C. Kochanek[16],
M. Makler[17], P. Martini[16], N. Morgan[18], B.Nord[2], F. Ostrovski[6,19], P. Schechter[20],
D. Tucker[2], R. Wechsler[21], T. M. C. Abbott[22], F. B. Abdalla[23,24] S. Allam[2],
A. Benoit-Lévy[25,26,23], E. Bertin[25,26], D. Brooks[23], D. L. Burke[12,27], A. Carnero Rosell[28,29],
M. Carrasco Kind[30,31], J. Carretero[32], M. Crocce[33], C. E. Cunha[12], C. B. D'Andrea[34],
L. N. da Costa[28,29], S. Desai[35], J. P. Dietrich[36,37], T. F. Eifler[38], B. Flaugher[2], P. Fosalba[14],
J. García-Bellido[39], E. Gaztanaga[14], M. S. Gill[27], D. A. Goldstein[40,41], D. Gruen[12,27],
R. A. Gruendl[30,31], J. Gschwend[28,29], G. Gutierrez[2], K. Honscheid[16,42], D. J. James[22,43],
K. Kuehn[44], N. Kuropatkin[2], T. S. Li[2,45] M. Lima[28,46], M. A. G. Maia[28,29], M. March[34]
J. L. Marshall[45], P. Melchior[47], F. Menanteau[30,31], R. Miquel[48,49], R. L. C. Ogando[28,29]
A. A. Plazas[38], A. K. Romer[50], E. Sanchez[51], ,R. Schindler[27] M. Schubnell[52],
I. Sevilla-Noarbe[51], M. Smith[53], R. C. Smith[22], F. Sobreira[28,54], E. Suchyta[55],
M. E. C. Swanson[31], G. Tarle[52], D. Thomas[9], A. R. Walker[22]

[1] *European Southern Observatory, Karl-Schwarzschild-Strasse 2, 85748 Garching bei München, DE*
[2] *Fermi National Accelerator Laboratory, Batavia, IL 60510*
[†] *Packard Fellow. The full list of affiliations can be found at the end of the paper.*

Accepted . Received

ABSTRACT

We present detailed modeling of the recently discovered, quadruply lensed quasar J0408-5354, with the aim of interpreting its remarkable configuration: besides three quasar images (A,B,D) around the main deflector (G1), a fourth image (C) is significantly reddened and dimmed by a perturber (G2) which is not detected in DES imaging data. From lens models incorporating (dust-corrected) flux-ratios, we find a perturber Einstein radius $0.04'' \lesssim R_{\rm E,G2} \lesssim 0.2''$ and enclosed mass $M_p(R_{\rm E,G2}) \lesssim 1.0\times10^{10}M_{\odot}$. The main deflector has stellar mass $\log_{10}(M_{\star}/M_{\odot}) = 11.49^{+0.46}_{-0.32}$, a projected mass $M_p(R_{\rm E,G1}) \approx 6\times10^{11}M_{\odot}$ within its Einstein radius $R_{\rm E,G1} = (1.85\pm0.15)''$ and predicted velocity dispersion $267 - 280$km/s. Follow-up images from a companion monitoring campaign show additional components, including a candidate second source between the quasar and G1. The predicted time-delays ($\Delta t_{AB} = (135.0\pm12.6)$d, $\Delta t_{BD} = (21.0\pm3.5)$d) are close to those measured, but better imaging is required for proper modeling and comparison.

Key words: gravitational lensing: strong – methods: statistical – astronomical data bases: catalogs – techniques: image processing

1 INTRODUCTION

Strongly lensed quasars are interesting astrophysical objects for diverse purposes (Courbin et al. 2002). The morphology of the multiple images, accompanied by arcs or rings tracing the lensed host galaxy, enables the description of the mass profile of the lens/deflector galaxy, which typically sits at redshifts $z_l \approx 0.5-1$ (e.g. Oguri et al. 2014). Thanks to magnification, the source can be super-resolved, well beyond what is possible for unlensed distant quasars. Astrometric and flux-ratio 'anomalies' among the multiple images are signatures of luminous and/or dark substructure surrounding the lens (Dalal & Kochanek 2002; Nierenberg et al. 2014), as well as faint features such as extended disks or isophotal twist, boxiness or diskiness (Möller et al. 2003; More et al. 2009; Vegetti et al. 2012; Hsueh et al. 2016; Gilman et al. 2016). When the source luminosity varies over time, the time delay between different images can be measured (e.g. Schechter et al. 1997; Tewes et al. 2013; Bonvin et al. 2016) and used to measure cosmological distances (as originally envisioned by Refsdal 1964, for lensed Supernovae) and hence the expansion rate of the Universe, yielding low-redshift (z_l) constraints on cosmological parameters that are independent of local distance-scale calibrations (cf. Treu & Marshall 2016; Suyu et al. 2016, and references therein).

Image-configuration has a central role for these studies. Systems with four images of the source quasar (hereafter *quads*) provide more information on the mass profiles of the deflector. In contrast, systems with two well-separated images (or *doubles*) can generally be more easily monitored for time variability with ground-based long-cadence observations, since fewer point sources must be de-blended within the same region. Systems in a *fold* configuration, where two of the quasar images are close to one another, are an interesting transition case, that allows for both robust time-delay measurements and lens mass reconstruction (Ding et al. 2016). In particular, in a fold configuration the source lies close to the caustic separating the double and quad regimes, with a *merging pair* of two of the images, thereby giving a highly stretched view of the quasar host near its centre (More et al. 2009; Rusu et al. 2014; Agnello et al. 2016).

Wide-field surveys offer a significant opportunity to discover new systems with suitable configuration, to be followed up for ancillary data. In particular, the Dark Energy Survey (hereafter DES: Sánchez & DES Collaboration 2010) has opened a new window for lens searches in the Southern Hemisphere, thanks to a combination of large footprint, depth and good image quality of the Dark Energy Camera (Flaugher et al. 2015; Dark Energy Survey Collaboration et al. 2016).

Here, we detail the first models of a new quasar lens, J0408-5354 (RA=62.091333, DEC=-53.900266). This lens was discovered by Lin et al. (2017) in the Y1A1 release of DES (Diehl et al. 2014; Drlica-Wagner et al. 2017), through a visual inspection of blue objects near red galaxies. Its multi-band images show four compact sources, compatible with being point-like given the DES point-spread-function (PSF), around a luminous red galaxy as shown in Figure 1. A spectroscopic confirmation campaign (Lin et al. 2017) shows that the three bright, blue point sources are images of the same source quasar at redshift $z_{\rm s}=2.375$, with absorption features at $z_{\rm l}=0.597$ that can be attributed to the lens galaxy. The fourth compact source to the South-West (fig. 1) is redder than the other confirmed quasar images. Detailed modeling is required to determine whether the anomalous colour is given by dust extinction, microlensing, or an additional red galaxy along the line of sight.

In this follow-up paper, we aim to shed light on the lensing nature of J0408-5354, expanding upon the discovery paper. First, we model the DES images to obtain object positions and spectral-energy distributions (SEDs). The multi-band SEDs of the point-sources can be used to quantify chromatic effects (such as microlensing or dust extinction), while the SED of the lens galaxy is used to estimate its stellar mass. The image positions are used as inputs to gravitational lens models, whose results are then used to estimate the dark matter content of the lens and verify whether an additional galaxy, lying very close to the reddened compact source along the line of sight, is needed to reproduce the observed flux ratios. We will show that based on the data available so far, the most plausible interpretation of the system consists of a main deflector galaxy and a satellite producing four images of a background lens quasar. The satellite deflector is very well aligned with one of the images, suppressing its flux and contaminating its colours.

This paper is structured as follows. In Section 2 we detail the multi-band model results of the DES *grizY* images. A comparison of different lens models is given in Section 3. We conclude in Section 4, including a discussion of the significance of J0408-5354 for different quasar lens searches, and briefly summarize in Section 5. Whenever needed, a standard flat ΛCDM cosmology is adopted with $\Omega_\Lambda = 0.7$ and $H_0 = 70$km/s/Mpc.

2 SYSTEM CONFIGURATION

J0408-5354 consists of point-like and extended objects (fig. 1), which are blended in the DES segmentation maps. In order to obtain robust SED measurements, in this Section we forward-model the *grizY* image cutouts as a superposition of objects, to recover robust magnitudes and relative positions with realistic uncertainties. The effective exposure time, as calculated by the reduction pipeline of single-epoch images is 540 s (resp. 225 s) for *griz* (resp Y) cutouts. The on-target[1] PSF FWHM ranges from $0.56''$ in $Y-$band to $0.59''$ in $g-$band. We deployed two independent versions of our own codes to model the cutouts; likelihood is explored via MCMC, whence the maximum-a-posteriori values and their uncertainties are obtained.

Follow-up imaging observations have been conducted with the Wide-Field Imager (WFI) on the 2.2m telescope in La Silla, to measure the time-delays between the light-curves of different images (Bonvin et al. 2017, in prep.). A coadd and optimal deconvolution (following Magain et al. 1998) of the best-seeing images obtained so far, shown in Figure 2, reveals a more complex structure: besides G1, A, B, G2/C and D, at least three additional 'blobs' are visible (G3,G4,G5), as well as a nearly complete Einstein ring with

---

[1] The DES data processing and calibration system is described by Bertin (2011), Mohr et al. (2012), and references therein; its implementation on the Y1A1 release is described by Drlica-Wagner, Sevilla, Rykoff et al. (2017, in prep).

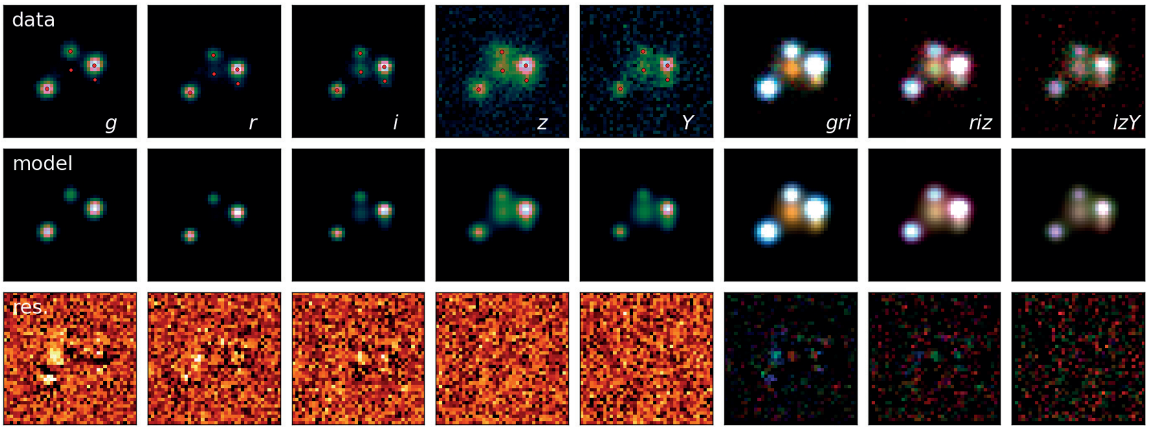


Figure 1. Multi-band images of J0408-5354 in *grizY*, from DES single-epoch data with best image quality, plus colour-composites (*gri*, *riz*, *izY*) in the last three columns. The data are shown in the first line with overlaid best-fit positions, the best-fit model (as detailed in Sect. 2) in the second line, and the residuals in the third line. An extra source between A and D is visible in the residuals, indicated as 'G3' in fig 2. Most of the residuals, besides G3, are due to PSF mismatch (around image A) and by blending of B and G2/C. North is up and East is left.

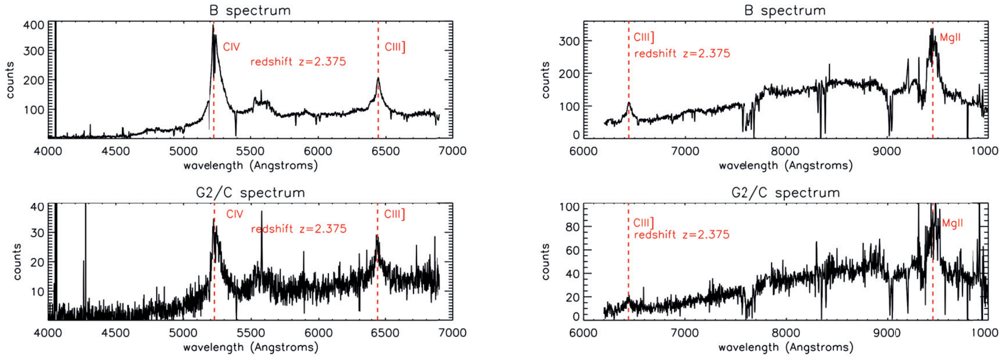


Figure 3. Follow-up spectra of images B and G2/C, obtained with a similar Gemini South setup as the discovery spectra of Lin et al. (2017). Both images have the same emission lines, but G2/C is considerably dimmed and reddened with respect to image B.

radius $\approx 1.6''$. Better data are needed to ascertain the nature of this ring and whether G3, G4, G5 are physically connected to it. The R_c-band image in Figure 2 has pixels of $0.12''$ per side and point-sources with a FWHM$= 0.2''$, allowing to locate the position angle (p.a.) of G1 to $\approx$ 30deg E of N. We will discuss these aspects further in the following Sections.

2.1 Image models

As illustrated in Figures 1 and 2, the system consists of a red galaxy (G1) surrounded by three blue point-like objects (A, B, D) and a redder and compact object (G2/C). As will be quantified in Section 3, if the system is a genuine quad, then G2/C would be a saddle-point image 'C', merging with B in a fold-like configuration.

Follow-up Gemini South spectra (program ID GS-2017A-LP-5, PI E. Buckley-Geer) were taken in March and April 2017, using the same setup as in Lin et al. (2017), except that B and G2/C were targeted on the same slit. This slit setup permits spatial decomposition of the spectra and resolves the previous issue of contamination of the G2/C spectrum by light from the much brighter image B. The results confirm that G2/C has the same quasar emission lines as B and the other blue images, albeit with strong differential reddening (Fig. 3).

Given the ordering of stationary points in the Fermat potential of a fold configuration (Saha & Williams 2003), the shortest arrival time corresponds to image A (minimum), followed by B (minimum) and C (first saddle-point) and then D (second saddle-point). For this reason, 'C' will be alternatively denoted as the *first saddle-point image* hereafter. Throughout this paper, we will treat this fourth image as an independent object, i.e. will not use its properties directly in constraining the lens models.

obj.	δRA($''$)	δDEC($''$)	g	r	i	z	Y
A	0.0	0.0	20.07±0.07	20.16±0.07	20.16±0.07	19.96±0.10	20.04±0.10
B	-6.34	1.85	19.98±0.07	19.95±0.07	19.74±0.10	19.28±0.08	19.34±0.10
G2/C	-6.43	0.75	22.68±0.20	21.98±0.15	21.46±0.15	20.91±0.12	20.56±0.16
D	-3.12	2.91	20.90±0.07	20.94±0.10	20.73±0.12	20.42±0.10	20.77±0.13
G1	-3.31	1.48	22.18±0.20	20.65±0.03	19.77±0.04	19.31±0.03	19.12±0.05
A	0.00	0.00	20.08±0.01	20.15±0.01	20.15±0.02	19.90±0.07	19.95±0.14
B	-6.35	1.86	19.86±0.01	19.79±0.01	19.66±0.02	19.29±0.07	19.25±0.15
G2/C	-6.42	0.69	23.16±0.11	21.61±0.05	20.92±0.06	20.82±0.09	20.45±0.10
D	-3.13	2.96	20.86±0.02	20.98±0.02	20.90±0.03	20.34±0.07	20.51±0.15
G1	-3.31	1.58	22.61±0.16	20.52±0.06	19.51±0.06	19.34±0.07	19.12±0.08
G3	-1.10	1.63	22.09±0.16	21.80±0.17	21.50±0.21	>21.20	>20.85

Table 1. Positions (relative to image A) and SEDs of the objects in J0408-5354, from a joint model of the DES $grizY$ single-epoch images with best image quality, adopting the DES-reconstructed PSF (*upper* sub-table) or a parametric fit to a nearby star (*lower* sub-table). Image A is at $(\mathrm{RA}, \mathrm{DEC}) = (62.091323, -53.900289)$. All the positions have an uncertainty of 0.25×10^{-4}deg $= 0.09''$, smaller than half the DES pixel size ($0.27''$), with zero covariance between δRA and δDEC. The naming scheme is illustrated in Figure 2. With the current depth and image quality, there are degeneracies in the fitted parameters of G3 and those of other components, primarily G1. The zY magnitudes of the 'blue plume' G3 are quoted as upper limits.

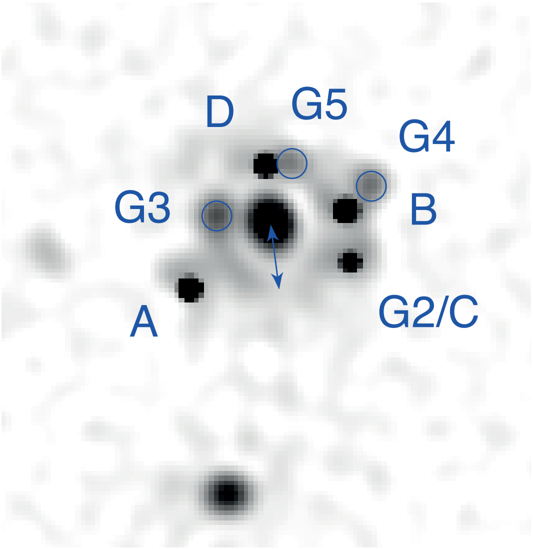


Figure 2. WFI R_c−band image of J0408-5354 after coadd and deconvolution of the best-seeing images. The central lens galaxy is $G1$. Image A, farther from $G1$, is the first to reach the observer, followed by image B, C and D; lens models will need a perturber G2 near the location of image C (see Sct. 3); three additional blobs are visible, marked by blue circles and denoted as G3,G4,G5. G3 sits on a nearly complete ring of radius $\approx 1.6''$, indicated by the blue arrow.

The DES cutouts are modelled as the superposition of a galaxy with a Sérsic (1968) profile for G1, and four point sources for A,B,D and G2/C. Different choices for the PSF are available, as it can be adopted from the DES PSF reconstruction or explicitly modelled as a superposition of analytic profiles. Each of these leads to a slight PSF mismatch on pixel-scales, but does not change the results appreciably. In order to test the robustness of the results, we opted for: (I) a model with the DES-reconstructed PSF; and (II) a model with a Moffat profile (Moffat 1969) fit to a nearby star to determine a parametric PSF. In the model, we impose that the relative displacements of all components (with respect to image A) are the same in every band. The model then comprises: the position angle ϕ_l, Sérsic index n_s and half-light radius R_{eff} of G1; the $grizY$ positions of A; the relative displacements of G1, B, G2/C and D; and the $grizY$ magnitudes of all objects. The Moffat PSF model (II) includes G3.

The inferred parameters with their uncertainties are listed in Table 1. Unfortunately, the depth and image quality of the survey cutouts are not sufficient to constrain n_s and R_{eff}. Nevertheless, the multi-band magnitudes of G1 (marginalized over everything else) are still well constrained. The (broad-band) SEDs of G1 and the four images are shown in Figure 4. The colours of image G2/C can be obtained by adding a standard reddening law (Cardelli et al. 1989, using $R_V = 3.1$ and $E(B-V) = 0.3$) to the SED of image B, but the overall magnitudes need an additional 'grey' dimming of 0.8mags; we also sum the small contribution of a putative galaxy G2 3.5mags fainter than G1, in order to better reproduce the $zY-$band fluxes. We will return to these points in Section 3.

2.2 Lens stellar mass

The $grizY$ SED inferred for the main deflector galaxy G1 can be used to estimate its stellar mass. We used the public[2] version of FAST (Kriek et al. 2009). Motivated by Treu et al. (2010), we adopt a Salpeter stellar IMF, which is expected for massive early-type galaxies. A direct measurement of the lens velocity dispersion would enable an IMF-independent determination of the stellar mass (Auger et al. 2009). The best-fit model is shown in Figure 4. With the uncertainties from the SED modelling, we obtain $\log_{10}(M_{\star}/M_{\odot}) = 11.49^{+0.46}_{-0.32}$. We will compare this to the results of lens models in the next Section.

---

[2] Available at `http://w.astro.berkeley.edu/~mariska/FAST.html`

	$\theta_{\mathrm{E},l}$	q	ϕ_l	γ_s	φ_s	b_p	s_p/b_p	$\theta_{\mathrm{E},p}$
SIE	$(1.98 \pm 0.08)''$	0.63 ± 0.06	-60.0 ± 2.0	—	—	—	—	—
SIS+XS	$(1.87 \pm 0.08)''$	[1.00]	—	0.13 ± 0.3	29.8 ± 3.3	–	—	—
SIS+XS+G2	$(1.73 \pm 0.15)''$	[1.00]	—	0.11 ± 0.03	18.4 ± 10.1	$(0.33 \pm 0.23)''$	0.24 ± 0.21	$(0.26 \pm 0.13)''$
SIS+XS+G2$^{(a)}$	$(1.72 \pm 0.10)''$	[1.00]	—	0.10 ± 0.02	16.5 ± 7.2	$(0.35 \pm 0.19)''$	0.30 ± 0.20	$(0.22 \pm 0.08)''$
SIE+XS+1p$^{(b)}$	$(1.90 \pm 0.10)''$	0.67 ± 0.04	-58.3 ± 1.4	0.016 ± 0.012	20.0 ± 3.0	$(0.051 \pm 0.008)''$	0.24 ± 0.17	$(0.033 \pm 0.013)''$ [C]
SIE+XS+2p$^{(b)}$	$(1.50 \pm 0.10)''$	0.93 ± 0.02	-59.4 ± 2.3	0.004 ± 0.032	19.2 ± 2.2	$(0.40 \pm 0.07)''$	0.14 ± 0.07	$(0.34 \pm 0.07)''$ [C]
						$(0.44 \pm 0.06)''$	0.05 ± 0.04	$(0.42 \pm 0.06)''$ [D]

Table 2. Inferred lens model parameters in the case of a Singular Isothermal Ellipsoid (SIE, first row), a Singular Isothermal Sphere plus external shear (SIS+XS, second row), or the same plus a small perturber G2 where G2/C is observed in the DES cutouts, adopting $0.2''$ for the positional uncertainties of A, B, D and $0.3''$ for those of G1 and G2. The lens p.a. of G1 (which may be different from that of its starlight) is quoted in 'mathematical notation' N of W, corresponding to ≈ 30 deg E of N. The perturber Einstein radius $\theta_{\mathrm{E},p}$ is not an additional parameter, being inferred directly from b_p and s_p. Models with a sub-critical G2 ($s_p > b_p/2$) are not excluded. $^{(a)}$This line shows the average parameters and standard deviations obtained when all uncertainties on positions are set to $0.1''$. $^{(b)}$ The two bottom lines list the lens parameters form a model with G1 plus one (upper) or two (lower) perturbers with free positions, and flux-ratios as additional constraints (Sect. 3.3.2). Letters in square brackets indicate the image next to which the perturbers lie.

model	$\log_{10}\mu(A)$	$\log_{10}\mu(B)$	$\log_{10}\mu(D)$	$\log_{10}\mu(C)$	$x_{\mathrm{C}} - x_{\mathrm{G1}}$ $('')$	$y_{\mathrm{C}} - y_{\mathrm{G1}}$ $('')$
SIE	0.47	0.89	0.52	0.76	1.65 ± 0.05	-0.89 ± 0.03
	0.45 ± 0.09	0.93 ± 0.17	0.45 ± 0.18	0.86 ± 0.20		
SIS+XS	0.52	1.05	0.58	1.14	1.60 ± 0.05	-0.70 ± 0.05
	0.64 ± 0.13	1.09 ± 0.15	0.71 ± 0.17	0.93 ± 0.19		
SIS+XS+G2	0.64	0.82	0.51	0.49	1.58 ± 0.03	-0.47 ± 0.07
	0.77 ± 0.19	0.97 ± 0.14	0.62 ± 0.16	0.70 ± 0.25		

Table 3. Inferred *logarithmic* magnifications for the three models fit to positions of A,B,D and G1, with one SIE (top) or SIS+XS (middle) in the lens plane, or with the addition of a perturber at G2/C (bottom). The first line of each block is $\log_{10}(\mu)$ from the best-fit model, while the second line shows the mean and standard deviation from the MCMC posterior. The last column lists the predicted displacement of image C, in terms of West-ward and North-ward displacements from the best-fitting position of $G1$ from Sect. 2 (identified with $\delta x = \delta y = 0$). The positional uncertainties are systematics-dominated, as the predicted position (especially y_{C}) can change appreciably across models.

3 LENS MODELS

The three images A, B, D have compatible SEDs, as is also confirmed by their long-slit spectra by Lin et al. (2017). We can then use their positions relative to G1 to model this system as a gravitational lens, obtaining estimates of the total mass (within the Einstein radius) and predicted time-delays between different images. Since G2/C is substantially redder than the other components, we do not include it in the lens model, but rather compare its properties with those predicted by the lens model fit to the other components. The technicalities of the lens model are described in Appendix A.

Conservatively, we adopt $0.2''$ positional uncertainties on A, B, D and $0.3''$ on G1, G2/C, about twice as large as those from the cutout modelling of Section 2 (relying solely on the DES cutouts). This allows us to explore a wide family of lens models and draw some general conclusions, in particular on the flux-ratios allowed by different models. In one case, we also allow the positional uncertainties to be those given directly by the cutout modelling (last line of tab. 2). The inferred lens model parameters for all models are given in Table 2. We stress that we are not using the smaller uncertainties from the WFI deconvolution, in order to highlight the robustness of some conclusions that held already with DES-quality data. However, when ellipticity is included in the lens model (defined as 'SIE' below), its p.a. agrees well with that from the WFI images shown in Figure 2.

The images A, B, D are mapped to the source plane according to the lens equation

$$\boldsymbol{\theta}_s = \boldsymbol{\theta}_{im} - \boldsymbol{\alpha} - \boldsymbol{\Gamma}\boldsymbol{\theta}_{im} \ , \tag{1}$$

where $\boldsymbol{\theta} = (\delta x, \delta y)$ is the angular displacement relative to the best-fitting G1 center from Section 2, the *external shear* matrix $\boldsymbol{\Gamma}$ is defined as

$$\boldsymbol{\Gamma} = \gamma_s \begin{pmatrix} \cos(2\varphi_s) & \sin(2\varphi_s) \\ \sin(2\varphi_s) & -\cos(2\varphi_s) \end{pmatrix} \tag{2}$$

and $\boldsymbol{\alpha}$ depends on how we describe the deflections by lensing galaxies. When describing lens galaxies, we use parametric models for their convergence profiles $\kappa = \Sigma/\Sigma_{cr}$, where $\Sigma_{cr} = \mathrm{c}^2 D_s/(4\pi \mathrm{G} D_l D_{ls})$ accounts for the dimensional dependence on angular-diameter distances. In particular, we use a Pseudo-Isothermal Ellipsoidal Mass Profile (PIEMD, Kassiola & Kovner 1993). This model provides a good representation of the gravitational potential of lens galaxies (e.g. Treu 2010) and the deflection angles $\boldsymbol{\alpha}$ in coordinates (X, Y)

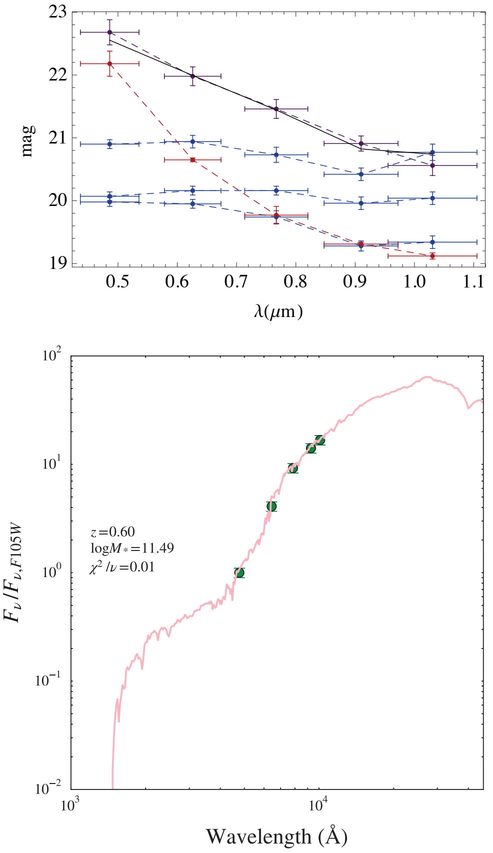


Figure 4. *Top: grizY* magnitudes of the multiple components; red (resp. blue) symbols indicate the galaxy G1 (resp. other compact images A,B,D), while the fainter SED with purple symbols corresponds to G2/C. The SED of image B, once reddened, needs an additional dimming of ≈ 0.8 mag in all bands to coincide with that of G2/C (black line), to which we also sum the contribution of a galaxy 3.5 mag fainter than G1 as discussed in Section 4. *Bottom:* Spectrum of the main deflector galaxy G1 from the best-fit FAST model, yielding $\log_{10}(M_\star/M_\odot) = 11.49^{+0.46}_{-0.32}$. The observed photometry is given by the dark-green symbols.

aligned with the principal axes of the iso-density ellipsoids

$$\alpha_X = -\frac{b}{\sqrt{1-q^2}}\arctan\left(\frac{X\sqrt{1-q^2}}{s+\sqrt{q^2(s^2+X^2)+Y^2}}\right) \quad (3)$$

$$\alpha_Y = -\frac{b}{\sqrt{1-q^2}}\operatorname{arctanh}\left(\frac{Y\sqrt{1-q^2}}{q^2 s+\sqrt{q^2(s^2+X^2)+Y^2}}\right) \quad (4)$$

are fully analytic, together with the convergence and the Fermat potential. The expression in coordinates (x, y) in West-North orientation requires just rotations in the coordinates and deflections, for which we choose the lens long-axis p.a. ϕ_l as positive N of W. The spherical $(q \to 1)$ and core-less $(s/b = 0)$ limit reduces to the Singular Isothermal Sphere (SIS), for which b is also the Einstein radius $R_{\rm E}$ enclosing a mean convergence of 1. In the Singular Isothermal Ellipsoid (SIE) case $(q < 1,\ s/b = 0)$, with the above notation we have $R_{\rm E} = b/\sqrt{q}$ as the ellipsoidal coordinate of the contour enclosing $\langle\kappa\rangle = 1$. In the case where $q = 1$ but $s/b > 0$, the Einstein radius is $R_{\rm E} = b\sqrt{1-2s/R_{\rm E}}$, which means that the PIEMD can be sub-critical ($\kappa < 1$ everywhere) when $s > b/2$. The Einstein radius can be used to estimate the lens velocity dispersion via[3]

$$\sigma_{\rm sis} = c\sqrt{\frac{R_{\rm E}D_s}{4\pi D_l D_{ls}}} = 203\,(\theta_{\rm E}/1'')^{1/2}{\rm km/s} \quad (5)$$

while the projected mass within R_E is

$$M_p(R_E) = \pi\Sigma_{\rm cr}R_E^2 = 2.0 * 10^{11}(\theta_{\rm E}/1'')^2 M_\odot\ , \quad (6)$$

regardless of the lens model. Here and in what follows, $\theta_E = R_E/D_l$ is the Einstein radius in angular units, the same as for the lens strength parameter b.

3.1 Models with one Deflector

For the first models, we describe the lensing mass distribution as given solely by G1. The first model (SIE) comprises simply a SIE representing G1. The second model (SIS+XS) adopts a SIS for G1, with the addition of external shear with non-null γ_s. The resulting parameters are listed in Table 2.

Both the SIE and SIS+XS models reproduce the positions of images A,B,D and predict a saddle-point image 'C' near the position G2/C found in Sect. 1 (fig. 6), whose relative position can vary from model to model, still within one or two DES pixels. The inferred Einstein radius $\theta_{{\rm E},l}$ of G1 is slightly less than half the A-to-B image separation ($\approx 2.2''$), due to quadrupole contributions to the deflection either by ellipticity or by shear. The quadrupole shear-ellipticity degeneracy is evident in that the shear angle φ_s in the SIS+XS case is orthogonal to the inferred lens position angle ϕ_l of the SIE case. The lens velocity dispersion and mass within $R_{\rm E}$ can be estimated as (286 ± 6)km/s and $(7.9 \pm 0.6)10^{11}M_\odot$ (resp. (280 ± 6)km/s and $(7.0 \pm 0.6)10^{11}M_\odot$) for the SIE (resp. SIS+XS) model.

Models with just one central deflector predict that image 'C' should be about as bright as image 'B', even with relatively large adopted uncertainties on the image positions ($0.2''$ instead of $0.09''$). This is summarized in Figures 5 and 7, and in Table 3.

3.2 Models with Perturbers

The first saddle-point predicted by models with one deflector would fall near the position of G2, which however is appreciably redder than the other images and significantly fainter than predicted even in the reddest bands. Extinction as measured in other lensed quasars (Dai et al. 2006; Mediavilla et al. 2005) does not differ substantially to that measured in the Milky Way and Magellanic Clouds (e.g. Cardelli et al. 1989). However, while the simple addition of a standard reddening law[4] to the SED of image B can reproduce the colours of image G2/C, it still requires a 'grey' dimming of ≈ 0.8mag in each band to match its overall magnitudes as in fig 5.

---

[3] The numerical prefactors in the second equalities are specific to the redshifts z_s, z_l of source and deflector in this particular case.
[4] With $R_V = 3.1$ and $E(B-V) = 0.3$, blueshifting the DES wavebands to the lens rest-frame.

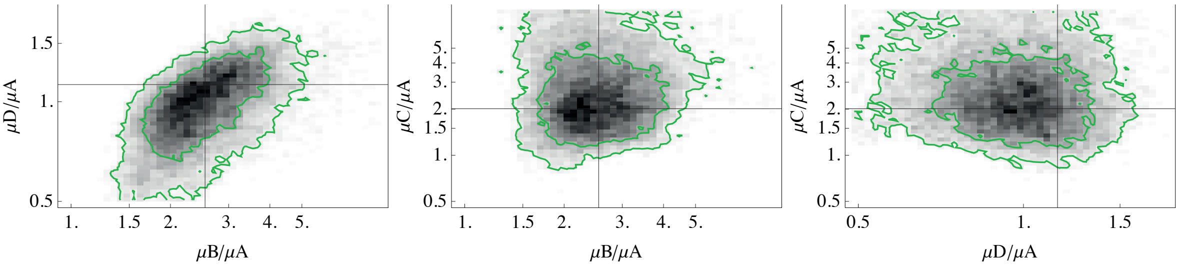


Figure 5. Output magnifications from a model with one SIE in the lens plane; the green contours represent the 68% and 95% quantiles of the marginalized posterior (no parameters held fixed). While the magnifications of B and D relative to A are in qualitative agreement with the SED fit results, the predicted image C should be almost as bright as image B and appreciably brighter than image A. This is not observed even after differential reddening is added to fit the colours (Sect. 2), and so it cannot be solely the result of dust extinction.

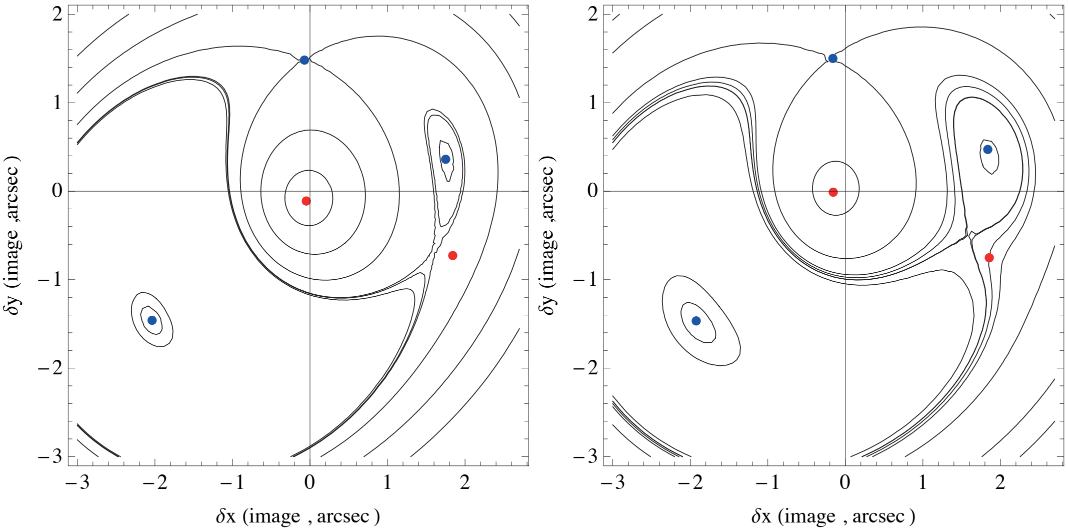


Figure 6. Time-delay contours for the case with one SIS plus external shear (*left* panel) or with the addition of a perturber at $G2$ (*right*). Models generally predict the fourth image position within one pixel-length in each direction from $G2$. This has a magnification comparable to that of image B if no perturber is present nearby. Being a saddle-point image, its magnification is easily suppressed by the presence of a small perturber at $G2$.

3.2.1 Fixed perturber at G2/C

Since G2/C lies close to image B, the differential reddening should be produced by a local overdensity, such as a small galaxy, whose lensing effect can also alter the magnification of image C. In general, saddle-points of the Fermat potential are *suppressed*, i.e. dimmed, by the presence of nearby perturbers, whereas minima fluctuate less (Schechter & Wambsganss 2002; Keeton 2003).

For this reason, we add a galaxy at the location of G2/C, which we describe as a PIEMD with $q = 1$. The addition of a perturber at a fixed position increases the number of parameters by two (core size and Einstein radius), making the model under-constrained. However, we can still rely on the priors on positions given by Section 2, and examine the range of parameter configurations that are compatible with the observed image configuration.

For simplicity, and due to the lack of an independent redshift measurement, we place the perturber in the same plane of the main lens G1. In general, models of lenses with four images have degeneracies among the monopole and quadrupole parameters (Kochanek et al. 2006). As verified above, the SIS+XS and SIE models do not differ appreciably in the output image positions and magnifications (tab. 2, 3, fig. 7).

The inferred lens parameters of the new model (with a main lens G1 and a perturber G2), given in Table 2, suggest a fairly small ($\approx 0.2''$) Einstein radius and do not rule out a sub-critical perturber. Even when $s_p < b_p/2$, the formation of multiple images and their magnifications depend on the distance between the perturber and image C; even though this does not occur in this configuration, the constraint that multiple images are not produced by the perturber (or not observed) could be used if sharper and deeper data are available. Similarly to the findings of Nierenberg et al. (2014) on a different lens, limits on how massive the perturber can be are given simply by the requirement that the other im-

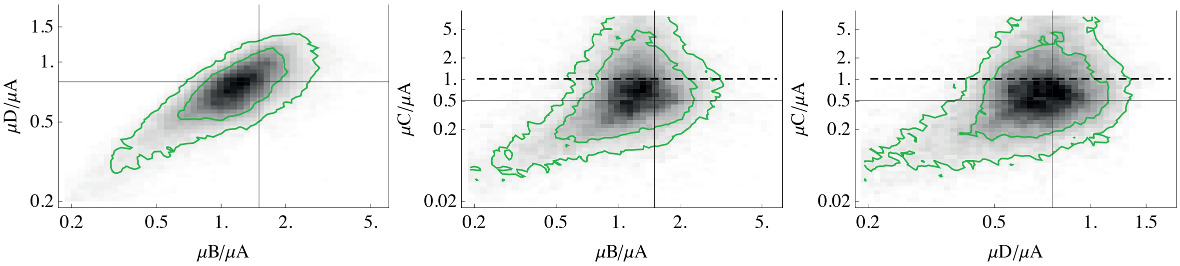


Figure 8. Output magnifications from a model with SIS+XS plus a small perturber (G2) at the observed location of G2/C; again, the green contours represent the 68% and 95% quantiles of the marginalized posterior. Whithin this class of models, both images *C* and *D* are slightly fainter than image *A* and significantly fainter than image *B*, even before dust extinction is accounted for. This saddle-point suppression is excessive on image C, due to the proximity of G2. For this reason, we explore models where one or two perturbers have have free position parameters, and flux ratios (corrected for dust and lightcurve delays) are used as constraints (see text, fig. 9).

ages (A,B,D) are not shifted by the perturber beyond their measured uncertainties.

With the addition of G2 in the lens model, the predicted image 'C' (fig. 6) is suppressed by the presence of the small perturber, making it slightly fainter than image A (fig. 8). With a small perturber at $\approx 0.2''$ from image C East-ward and North-ward, its SED can be easily reddened even though it lies very close to B. The small separation between C and G2 makes them hardly distinguishable even in the Gemini acquisition image of Lin et al. (2017), whose PSF has a quoted FWHM$\approx 0.5''$. Within this model, the lens velocity dispersion of G1 is (267 ± 12)km/s, and its projected mass within R_{E} is $(6.0 \pm 1.0)10^{11} M_{\odot}$. Even though G2 is not excluded to be sub-critical, we can still estimate its velocity dispersion and enclosed projected mass as (95 ± 17)km/s and $\lesssim 1.0 \times 10^{10} M_{\odot}$, respectively.

model	Δt_{AB} (days)	Δt_{BD} (days)	Δt_{BC} (days)
SIE	125	20	8
SIS+XS	100	19	2.5
SIS+XS+G2	87	29	6
SIE+XS+1p	135.0±12.5	21.0±3.5	1.8±0.3
SIE+XS+2p	140.0±12.0	14.0±2.4	0.0±2.0

Table 4. Time-delays predicted by different lens models, adopting flat-ΛCDM cosmology with $H_0 = 70$km/s/Mpc. Models using $0.2''$ positional errors have large time-delay uncertainties and posteriors that are very skewed from those of the best-fit solution. The delays from models with free perturbers are in good agreement with those subsequently measured by Bonvin et al. (2017 in prep.), but sharper and deeper imaging data are needed for proper modeling and comparison.

3.2.2 Perturbers with free positions

A perturber (G2) at the location of G2/C, with constraints set only by the relative image positions, is even too effective at saddle-point suppression. In fact, the magnifications predicted by that model (tab. 3) are to be meant *before* dust-extinction, which in this case is substantial, as mentioned in Section 2. Then, in order to properly assess the presence of a perturber, one must correct the magnitudes from Section 2 for dust-extinction, microlensing and delays in the lightcurves, and use those as constraints to the lens models. The long-slit spectra of (Lin et al. 2017) show some microlensing, in that flux ratios among emission lines are different from flux-ratios among the continua, but the amount is small ($\approx 10\%$) and, to our aim, negligible in the wavelength range $6000\mathring{A} < \lambda < 10000\mathring{A}$. Differential reddening can be inferred quite robustly from the differences in colours among different images, as mentioned above. In order to correct for variability, one needs shifted lightcurves from a monitoring campaign. In what follows, we use shifted magnitudes from the 2.2m-WFI campaign (Bonvin et al. 2917, in prep.), correct them for differential extinction and neglect microlensing effects. The grey dimming on image C is robust since its delay from image B is small.

Here, we analyse two models with perturbers with free positions, Einstein radii and core radii. In the first case, one perturber near the fold pair (B,C) is added to the primary deflector (G1), and a magnitude difference $m_C - m_B = 2.5\log_{10}(\mu_B/\mu_C) = (0.85 \pm 0.1)$ is added as constraint. In the second case, a second perturber is added near image D and the dust-corrected, delay corrected magnitude differences $m_D - m_A = (0.45 \pm 0.05)$, $m_A - m_B = (1.95 \pm 0.05)$ (Bonvin et al. 2017, in prep.) are additionally used as constraints. The model with two perturbers is explored because a model with just one perturber predicts higher magnifications for image D than are observed. As a model for G1, we adopt a SIE+XS with the following priors: uniform in $1.3'' < b_l < 2.2''$, $0.0 < \gamma_s < 0.14$, $0.5 < q < 0.99$; Gaussian in $\phi_l = (-60.0 \pm 5.0)$ deg, $\phi_s = (20.0 \pm 5.0)$ deg. Uniform priors are used for the parameters of the perturbers. The inferred parameters are appended in the two bottom lines of Table 2, and positions sampled from the posteriors are displayed in Figure 9. The 'butterfly' pattern in the allowed perturber positions is common to other lenses (e.g. Nierenberg et al. 2014).

3.3 Time-delays

From the lens models, we can also give some forecasts on the expected delays between the arrival times

$$t_i = \frac{(1+z_l)D_l D_s}{cD_{ls}}\left[\frac{1}{2}\left|\boldsymbol{\theta}_{im,i} - \boldsymbol{\theta}_s\right|^2 - \Phi\right] , \qquad (7)$$

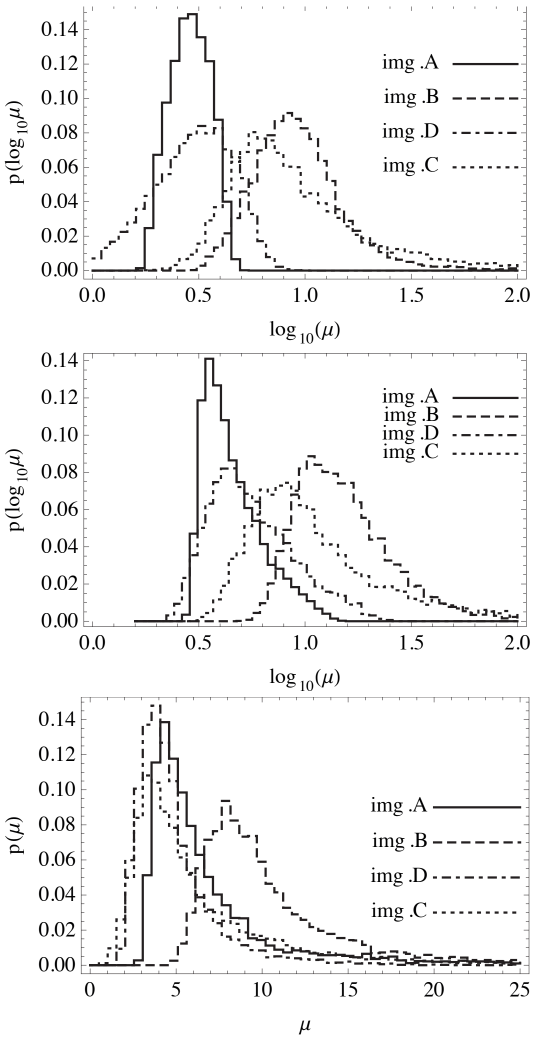


Figure 7. Output magnifications from a model SIE (top), SIS+XS (middle), and SIS+XS+G2 (bottom, see text), in logarithmic units. The SIE and SIS+XS models produce similar results, particularly for the predicted ordering of magnifications. Adding a perturber (G2) at the observed location of G2/C dims image C considerably.

where the projected potential Φ is analytic in all models chosen.

Their values are approximately the same across different models, albeit with $\approx$ 20day differences, and the ordering is always the same: the first image is A, followed by B, C shortly after, and finally D. The ordering is general and does not depend on whether a perturber is included in the model, being determined by the configuration of critical points (e.g. Saha & Williams 2003). Within the SIS+XS+G2 model, we have $\Delta t(AB) = 85$d, $\Delta t(BC) = 6$d, and $\Delta t(BD) = 29$d, where $\Delta t(i, j) = t_j - t_i$ is positive when the arrival-time of image i is shorter than that of image j. The quoted values have large uncertainties, due to the wide degeneracies in the lens models, and their marginalized posterior is offset from the values from the best-fitting model.Models with free perturbers have a more symmetric posterior, and smaller uncertainties in the predicted time-delays. Delays like these are ideal for ground based monitoring, because they are long enough to yield 1-3% precision with daily cadence, yet short enough that one or two observing seasons are sufficient. The delays predicted by models with free perturbers are in good agreement with those accurately measured by an ongoing, high-cadence 2.2m-WFI monitoring campaign (Bonvin et al. 2017, in prep.). However, the current data and models are not sufficient to allow for a proper comparison, primarily because of long-term microlensing effects and uncertainties in the lens potential.

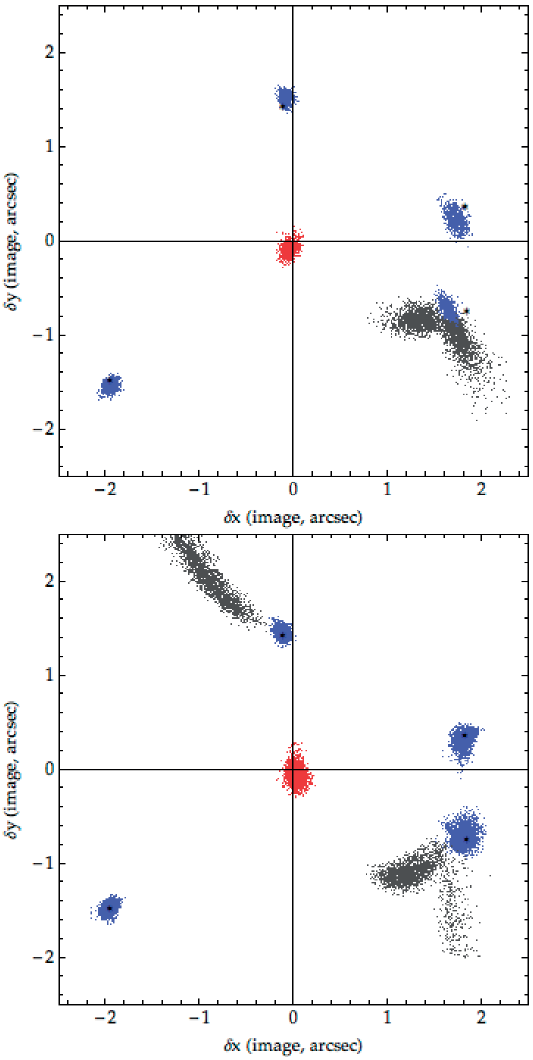


Figure 9. Poisitions of G1 (red), perturbers (grey) and model-predicted quasar images (blue) as sampled from the posteriors with one (*top*) or two (*bottom*) perturbers with free positions. The black star-symbols mark the image positions (relative to G1) from Section 2.

4 DISCUSSION

We have modelled J0408-5354 to obtain the photometry of its individual components, the stellar mass of the main lens galaxy and lens parameters for a choice of plausible models. The predicted time-delays and image configuration make this system amenable to follow-up for time-delay cosmography, as well as for studies of the quasar host near the central engine and substructure near the quasar images.

With the current data quality, there are vast degeneracies in the lens model parameters, which however can be easily relieved with high-resolution imaging data. This will also help locate the pertuber G2 responsible for the reddening *and* dimming of image C. The occurrence of both cases would not be uncommon, as seen e.g. for the lens B1608+656 (Myers et al. 1995; Fassnacht et al. 1996; Suyu et al. 2009). A direct measurement of the lens velocity dispersion, together with a follow-up campaign for time-delays, would yield a direct measurement of the angular-diameter distance to the lens via $D_l \propto \mathrm{c}^3\Delta t/\sigma^2$ (see Paraficz & Hjorth 2009; Jee et al. 2016, for a general discussion).

4.1 System Configuration

J0408-5354 consists of three blue point-like images of the source quasar at $z_s = 2.375$, and two redder components of which G1, the main lens, is at $z_l = 0.597$, whereas the nature of G2/C is unclear, given its colours and the significant contamination from spectra of other components (Lin et al. 2016). We have modelled this system as a superposition of an extended galaxy (G1) plus four compact sources A, B, D, G2/C and obtained deconvolved SEDs. In particular, the SED of G1 suggests a stellar mass $M_\star \approx 3.2\times10^{11}M_\odot$ for the lens, within ≈ 0.4dex uncertainty. There is a degree of systematic uncertainty in the positions and fluxes of B and G2/C, given their proximity and the fact that B is more than a magnitude brighter than G2/C in bluer bands. Some faint residuals are given by PSF mismatch on pixel scales, regardless of the choice of PSF. The WFI images with best seeing, once deconvolved, show what could be an additional ring with radius $R \approx 1.6''$, which is appreciably smaller than that inferred from lens models based on the positions of images A,B,D (as summarized below). If it corresponds to a second source, it can map to a unique location just if it sits between the quasar and the deflector.

4.2 Lens Model Properties

The positions of images A, B, D relative to G1 have been used to explore lens models of J0408-5354. Models with one main lens, adopted as Singular Isothermal Ellipsoid or Singular Isothermal Sphere plus external shear, predict an Einstein radius $\approx 1.9''$ and a lens p.a.≈ -60deg North of West, or 0.1 shear almost orthogonal to it. Both models, while successful at reproducing the positions of images A, B, D, would predict a saddle-point image where G2/C lies and about as bright as image B, which is not observed even in band Y or after differential extinction corrections. Models with a small perturber at the location of G2 reproduce the same image positions, but are able to suppress image C by about a magnitude even before dust is accounted for.The dust-corrected and delay-corrected flux ratios, as well as the images measured from the DES cutouts, are well reproduced by two small perturbers near the saddle-points.

The projected mass within the Einstein radius is $M_p(R_\mathrm{E}) \approx 6\times10^{11}M_\odot$, about twice the stellar mass estimated from the SED of G1. A proper evaluation of the dark matter fraction in the lens, however, would require a measurement of the effective radius of G1. When the perturber has non-null Einstein radius, its enclosed mass is $M_p(G2) \approx 1.3\times10^{10}M_\star$. The contribution of a small galaxy with magnitudes $m(G2) = m(G1) - 2.5\log_{10}(M_p(G2)/M_p(G1))$ is barely noticeable in *gri* bands, which in turn can be well reproduced by reddening and offsetting the SED of image B by ≈ 0.8 mag, and makes the $zY-$band magnitudes of image G2/C in complete agreement with the values measured from Section 2 (black line in fig. 4).

The estimated time-delays between images A,B,D (from lens models with free perturbers) are in agreement with what is measured by Bonvin et al. (2017, in prep.) through a 2.2m-WFI dedicated monitoring campaign. Still, given the uncertainties on image positions and few constraints, the derived uncertainties are appreciable and higher-resolution imaging data will be required to tighten the model-predicted uncertainties on the delays.

If indeed two sources are present at different redshift, J0408-5354 can also be used to measure Dark Energy cosmological parameters via the ratio of distance ratios D_s/D_{ls} to the different sources (e.g. Paczynski & Gorski 1981; Soucail et al. 2004; Collett et al. 2012), besides time-delay cosmography to measure H_0. The only other system with time delays and multiple source-planes that is known and studied to date is the galaxy cluster MACSJ1149.5+2223 (Treu et al. 2016).

4.3 Relevance of J0408-5354 for lens searches

The photometry and configuration of J0408-5354 make it an interesting testbed for different techniques of lensed quasar candidate selection. These, in turn, have implications for substructure studies, as the composition of lens-selected or source-selected samples affects the sensitivity to substructure, especially for lens searches that are tailored on simple lenses or on systems dominated by 'isolated' quasar SEDs.

Like the serendipitous quad of More et al. (2016), J0408-5354 was originally found by visual inspection of objects selected solely on *gri* survey properties, instead of relying on hybrid infrared 'excess' colours (Warren et al. 2000) that have been used to target quasars (Maddox & Hewett 2006; Maddox et al. 2012; Peters et al. 2015) or lenses (Ofek et al. 2007) and applied in other lens searches in DES (Agnello et al. 2015a; Ostrovski et al. 2017). After the initial discovery via the blue-near-red search of Lin et al. (2017), different teams have examined their own search methods. Here we provide a summary of the different findings.

4.3.1 Cutout classification: CHITAH

CHITAH (Chan et al. 2015) examines the image cutouts of objects to detect at least two blue compact sources and a red galaxy, evaluating how plausible the configuration is as a strong lens via the corresponding source-plane χ^2. This approach relies on the requirement that the blue images have very similar SEDs, distinct from the lens SED.

When applied to the *grizY* cutouts of J0408-5354, it did not flag this system as a possible quad since the fourth image G2/C is significantly redder than the others. However, based on A, B and G1, it did classify this system as a possible double. These findings suggest that pixel-based automatic recognition, such as CHITAH or LensTractor[5] could be made more flexible by accounting for possible SED variations of the predicted images.

4.3.2 Target selection: data mining

The first technique used to select lensed quasars in the DES relied upon Artificial Neural Networks (ANNs) trained on SDSS *griz* and WISE (Wright et al. 2010) $W1, W2$ bands of four main classes of objects (Agnello et al. 2015a). Despite the success of the first discovery results (Agnello et al. 2015b), further improvements could be made for wider application to DES, as discussed in Appendix B. With these new ANNs, J0408-5354 was automatically flagged as an extended quasar with $z_s > 1.75$, one of the two classes (besides 'lens') to be retained for visual inspection[6]. Despite the improvement in the ANNs and the *blind* re-discovery of J0408-5354, there is considerable scatter in the SDSS-DES translated magnitudes, which can cause some interesting objects to slip out of the selection boundaries (and false positives to leak in). The outlier selection method (Agnello 2017), in which J0408-5354 is rediscovered as a $> 3\sigma$ outlier among quasars and with low probability to be a galaxy, is somewhat immune from this issue, as are Population Mixture classifications (Ostrovski et al. 2017; Williams et al. 2016).

5 SUMMARY

J0408-5354 has an interesting fold-like image configuration, with three well-separated images (A,B,D) and a fourth one (C) in a merging pair with the brightest image (B). Besides the three, clearly identifiable blue images of the source quasar, a fourth component G2/C is fainter than simple lens-model predictions and appreciably red. While image B is already redder than the farthest image A, with $\Delta(Y - r) \approx 0.65$ compatible with a simple (Cardelli et al. 1989) reddening law with $E(B - V) = 0.3$, image G2/C is further reddened (additional $E(B - V) = 0.3$) and also requires a grey dimming of ≈ 0.8 mag in every band.

A small perturber ($R_{\mathrm{E,p}} \approx 0.2''$, $M_p \approx 1.0 \times 10^{10} M_{\odot}$) near the location of G2/C explains both the needed reddening and dimming over the whole *grizY* range. The data on positions and (dust-corrected, delay-corrected) flux ratios allow for two perturbers near the saddle-point images C and D, albeit with very uncertain masses (tab. 2) due to the current scarcity of constraints.

The image separation makes this system particularly apt to time-delay measurements. The B-A delay $\approx 120 - 140$days predicted by the full models (SIE+XS+pert.) is in good agreement with the results of a dedicated 2.2m-WFI monitoring campaign (Bonvin et al 2017, in prep.). The lens mass within the Einstein radius $R_{\mathrm{E}} = 1.73''$ is $M_p \approx (6.0 \pm 1.0) \times 10^{11} M_{\odot}$, about twice the stellar mass $M_{\star} \approx 3 \times 10^{11} M_{\odot}$ of the main galaxy G1.

The chromaticity and morphology of J0408-5354 mean that different search techniques, while successfully flagging it as a lens candidate, are triggered by different features. Also, the peculiar colours and configuration of the quasar images are a powerful reminder that automated search techniques should be flexible enough to encompass these systems, in view of homogeneous lens-selected or source-selected samples for follow-up science. Oguri & Marshall (2010) estimated 1146 quasar lenses within a depth of $i = 23.6$ in the 5000deg^2 final DES footprint, of which 14% are quads. Past and ongoing lens searches show that a suite of complementary techniques are needed to maximize the number of detected lenses, especially at magnitudes fainter than $i \approx 19$.

The composition of J0408-5354, with a primary (massive) lens and a small perturbers and a merging image-pair, make it both an interesting system for follow-up and a rather peculiar system to model. Spectroscopic and high-resolution imaging observations would enable more accurate models, both for cosmography and for substructure studies, and a highly magnified view of the source quasar and its host.

ACKNOWLEDGMENTS

AA and TT acknowledge support by the Packard Foundations through a Packard Research Fellowship and by the National Science Foundation through grant AST-1450141.

This paper was written as part of the STRong lensing Insights into the Dark Energy Survey (STRIDES) collaboration, a broad external collaboration of the Dark Energy Survey, `http://strides.astro.ucla.edu`

We are grateful to the anonymous referee for a 'fresh' and thorough reading of the paper, which helped clarify several points.

Funding for the DES Projects has been provided by the DOE and NSF(USA), MISE(Spain), STFC(UK), HEFCE(UK). NCSA(UIUC), KICP(U. Chicago), CCAPP(Ohio State), MIFPA(Texas A&M), CNPQ, FAPERJ, FINEP (Brazil), MINECO(Spain), DFG(Germany) and the Collaborating Institutions in the Dark Energy Survey. The Collaborating Institutions are Argonne Lab, UC Santa Cruz, University of Cambridge, CIEMAT-Madrid, University of Chicago, University College London, DES-Brazil Consortium, University of Edinburgh, ETH Zürich, Fermilab, University of Illinois, ICE (IEEC-CSIC), IFAE Barcelona, Lawrence Berkeley Lab, LMU München and the associated Excellence Cluster Universe, University of Michigan, NOAO, University of Nottingham, Ohio State University, University of Pennsylvania, University of Portsmouth, SLAC National Lab, Stanford University, University of Sussex, and Texas A&M University. The DES Data Management System is supported by the NSF under Grant Number AST-1138766. The DES participants from Spanish institutions are partially supported by MINECO under grants AYA2012-39559, ESP2013-48274, FPA2013-47986, and Centro de Excelencia Severo Ochoa SEV-2012-0234. Research leading to these results has received funding from the ERC under the EU's

---

[5] Available at `https://github.com/davidwhogg/LensTractor`

[6] In particular, the blend D+G1 with catalogue ID=3070264166, RA=62.0904688061, DEC=-53.8996413857

7th Framework Programme including grants ERC 240672, 291329 and 306478.

REFERENCES

Agnello, A., Kelly, B. C., Treu, T., & Marshall, P. J. 2015, MNRAS, 448, 1446

Agnello, A., Treu, T., Ostrovski, F., et al. 2015, MNRAS, 454, 1260

Agnello, A., Sonnenfeld, A., Suyu, S. H., et al. 2016, MNRAS, 458, 3830

Agnello, A., 2017, MNRASsubm., arXiv:1705.08900

Assef, R. J., Stern, D., Kochanek, C. S., et al. 2013, ApJ, 772, 26

Auger, M. W., Treu, T., Bolton, A. S., et al. 2009, ApJ, 705, 1099

Bertin, E. 2011, Astronomical Data Analysis Software and Systems XX, 442, 435

Bonvin, V., Tewes, M., Courbin, F., et al. 2016, AA, 585, A88

Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245

Chan, J. H. H., Suyu, S. H., Chiueh, T., et al. 2015, ApJ, 807, 138

Collett, T. E., Auger, M. W., Belokurov, V., Marshall, P. J., & Hall, A. C. 2012, MNRAS, 424, 2864

Courbin, F., Saha, P., & Schechter, P. L. 2002, Gravitational Lensing: An Astrophysical Tool, 608, 1

Dai, X., Kochanek, C. S., Chartas, G., & Mathur, S. 2006, ApJ, 637, 53

Dalal, N., & Kochanek, C. S. 2002, ApJ, 572, 25

Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, MNRAS, 460, 1270

Diehl, H. T., Abbott, T. M. C., Annis, J., et al. 2014, Proc. SPIE, 9149, 91490V

Ding, X.-H., et al. (2016) MNRASsubm.

Drlica-Wagner, A., & DES Collabroation 2017, in prep.

Fassnacht, C. D., Womble, D. S., Neugebauer, G., et al. 1996, ApJL, 460, L103

Gilman, D., Agnello, A., Treu, T., Keeton, C. R., & Nierenberg, A. M. 2017, MNRAS, 467, 3970

Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, AJ, 150, 150

Hsueh, J.-W., Fassnacht, C. D., Vegetti, S., et al. 2016, arXiv:1601.01671

Inada, N., Oguri, M., Shin, M.-S., et al. 2012, AJ, 143, 119

Jee, I., Komatsu, E., Suyu, S. H., & Huterer, D. 2016, JCAP, 4, 031

Kassiola, A., & Kovner, I. 1993, Liege International Astrophysical Colloquia, 31, 571

Keeton, C. R. 2003, ApJ, 584, 664

Kochanek, C.S., Schneider, P., Wambsganss, J., 2004, Part 2 of Gravitational Lensing: Strong, Weak & Micro, Proceedings of the 33rd Saas-Fee Advanced Course, G. Meylan, P. Jetzer & P. North, eds. (Springer-Verlag: Berlin)

Koopmans, L. V. E., Garrett, M. A., Blandford, R. D., et al. 2002, MNRAS, 334, 39

Kriek, M., van Dokkum, P. G., Labbé, I., et al. 2009, ApJ, 700, 221

Lin, H., & DES Collaboration (2017) subm.

Maddox, N., & Hewett, P. C. 2006, MNRAS, 367, 717

Maddox, N., Hewett, P. C., Péroux, C., Nestor, D. B., & Wisotzki, L. 2012, MNRAS, 424, 2876

Magain, P., Courbin, F., & Sohy, S. 1998, ApJ, 494, 472

Mediavilla, E., Muñoz, J. A., Kochanek, C. S., et al. 2005, ApJ, 619, 749

Moffat, A. F. J. 1969, AA, 3, 455

Mohr, J. J., Armstrong, R., Bertin, E., et al. 2012, Proc. SPIE, 8451, 84510D

More, A., McKean, J. P., More, S., et al. 2009, MNRAS, 394, 174

More, A., Lee, C.-H., Oguri, M., et al. 2016, arXiv:1608.06288

Möller, O., Hewett, P., & Blain, A. W. 2003, MNRAS, 345, 1

Myers, S. T., Fassnacht, C. D., Djorgovski, S. G., et al. 1995, ApJL, 447, L5

Nierenberg, A. M., Treu, T., Wright, S. A., Fassnacht, C. D., & Auger, M. W. 2014, MNRAS, 442, 2434

Ofek, E. O., Oguri, M., Jackson, N., Inada, N., & Kayo, I. 2007, MNRAS, 382, 412

Oguri, M. 2010, PASJ, 62, 1017

Oguri, M., & Marshall, P. J. 2010, MNRAS, 405, 2579

Oguri, M., Rusu, C. E., & Falco, E. E. 2014, MNRAS, 439, 2494

Ostrovski, F., McMahon, R. G., Connolly, A. J., et al. 2017, MNRAS, 465, 4325

Paczynski, B., & Gorski, K. 1981, ApJL, 248, L101

Paraficz, D., & Hjorth, J. 2009, AA, 507, L49

Peters, C. M., Richards, G. T., Myers, A. D., et al. 2015, ApJ, 811, 95

Refsdal, S. 1964, MNRAS, 128, 307

Rusu, C. E., Oguri, M., Minowa, Y., et al. 2014, MNRAS, 444, 2561

Saha, P., & Williams, L. L. R. 2003, AJ, 125, 2769

Sánchez, E., & DES Collaboration 2010, Journal of Physics Conference Series, 259, 012080

Schechter, P. L., Bailyn, C. D., Barr, R., et al. 1997, ApJL, 475, L85

Schechter, P. L., & Wambsganss, J. 2002, ApJ, 580, 685

Sersic, J. L. 1968, Cordoba, Argentina: Observatorio Astronomico, 1968,

Soucail, G., Kneib, J.-P., & Golse, G. 2004, AA, 417, L33

Suyu, S. H., Marshall, P. J., Blandford, R. D., et al. 2009, ApJ, 691, 277

Suyu, S. H., Bonvin, V., Courbin, F., et al. 2016, arXiv:1607.00017

Tewes, M., Courbin, F., & Meylan, G. 2013, AA, 553, A120

Treu, T., Auger, M. W., Koopmans, L. V. E., et al. 2010, ApJ, 709, 1195

Treu, T. & Marshall, P.J. 2016, Astron. Astrophys. Rev. 24: 11. doi:10.1007/s00159-016-0096-8

Treu, T., Brammer, G., Diego, J. M., et al. 2016, ApJ, 817, 60

Vegetti, S., Lagattuta, D. J., McKean, J. P., et al. 2012, Nat, 481, 341

Warren, S. J., Hewett, P. C., & Foltz, C. B. 2000, MNRAS, 312, 827

Williams, P., Agnello, A., & Treu, T. 2017, MNRAS in press, arXiv:1612.03821

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868-1881

APPENDIX A: LENS MODELING SPECIFICS

Regardless of the model specifics, all images must map to the same source-position. For each choice of the lens model parameters, a source at $\boldsymbol{\theta}_s$ in the source plane corresponds to images $\boldsymbol{\theta}_i$ in the image plane, and the goodness-of-fit can be described by the image-plane χ^2

$$\chi^2_{ip} = \sum_{i=1}^{3} \frac{|\boldsymbol{\theta}_i - \boldsymbol{\theta}_{im,i}|^2}{\delta_i^2} = \sum_{i=1}^{3} \frac{|\mathbf{A}_i(\boldsymbol{\theta}_s - \boldsymbol{\theta}_{s,i})|^2}{\delta_i^2}\ , \qquad \text{(A1)}$$

where $\boldsymbol{\theta}_{im,i}$ and $\boldsymbol{\theta}_{s,i}$ are the measured image-positions and their model-predicted source-plane positions for images A,B,D, $\mathbf{A}_i = \partial\boldsymbol{\theta}_{im,i}/\partial\boldsymbol{\theta}_{s,i}$ is the magnification tensor around each image and δ_i is the positional uncertainty on image i. The second equality relies on the fact that, near a reasonable lens solution, we can linearize the lens equation around the measured image positions. Then, within this approach, the best-fit source-position and corresponding image-plane χ^2 and image-positions can be found by straightforward linear operations. Its validity has been tested extensively by Oguri (2010). Based on those functional tests and ours, the lens models are accurate provided one iterates ≈ 30 times between the linearized best-fit $\boldsymbol{\theta}_s$ and $\boldsymbol{\theta}_{im,i}$, ensuring that all model image-positions are remapped to the same source-point.

Writing the χ^2 as above relies on a Gaussian distribution of the measured image positions, with isotropic positional uncertainties, and is equivalent to drawing image positions with infinite precision from Gaussians $\mathcal{G}(\boldsymbol{\theta}_{im,i}, \delta_i)$, considering (for each choice) a highly-penalized image-plane $\chi^2 = p\chi^2_{ip,1}$ in the lens model with

$$\chi^2_{ip,1} = \sum_{i=1}^{3} |\boldsymbol{\theta}_i - \boldsymbol{\theta}_{im,i}|^2 \qquad \text{(A2)}$$

and $p \gg \delta_i^{-2}$. This allows us to generalize the lens model likelihood to image configurations that do not have isotropic and Gaussian uncertainties. In particular, we can draw the relative displacements of G1, B and D with respect to image A as given by the likelihood explored in Section 2, which we call $\mathcal{L}_{SED}$. At very high values of p, the only parameter combinations that are explored are those that correspond to all image positions mapping back to the same source position, because other configurations are heavily penalized.

Another hypothesis underlying this approach is that the measured image position uncertainties are simply given by the extraction of Section 2, so that each image carries a weight proportional to its (squared) magnification in the χ^2. This does not account for systematic uncertainties in the image positions given by the proximity of different objects and PSF mismatch. This problem is evident for the brightest image B, which would instead carry the highest weight in χ^2_{im}. We then opt for a penalized source-plane χ^2 of the form

$$\chi^2_{sp} \;=\; p\sum_{j=1}^{3} |\boldsymbol{\theta}_{s,j} - \langle\boldsymbol{\theta}_s\rangle|^2\ , \qquad \text{(A3)}$$

where $\langle\boldsymbol{\theta}_s\rangle = (\boldsymbol{\theta}_{s,A} + \boldsymbol{\theta}_{s,B} + \boldsymbol{\theta}_{s,C})/3$ for each choice of the model parameters, and consider the lens-model likelihood as

$$\mathcal{L} \;\propto\; \mathcal{L}_{SED}(\boldsymbol{\theta}) \times \mathrm{e}^{-\chi^2_{\rm sp}/2}\ . \qquad \text{(A4)}$$

The penalty parameter p is gradually increased, until all possible models are effectively producing images originating from the same source-position, within milli-arcsecond tolerance, and the model uncertainties are driven by $\mathcal{L}_{SED}$.

APPENDIX B: MINING ACROSS SURVEYS

The original implementation of ANNs was based upon SDSS data and four main classes of objects. In order to be more widely applicable to DES, it was improved in three ways: (I) more object classes, including multiple redshift intervals for the 'quasar' class to distinguish low-redshift contaminants from higher-redshift objects; (II) less restrictive colour-cuts, that would otherwise exclude known lenses with higher $g-i$ or lower $W1-W2$; and (III) accounting for the differences in photometry between SDSS and DES via a cross-calibration valid for blue extended objects[7]. The best-fit regressions have

$$\begin{aligned} g_{\rm des} &= g_{\rm sdss} + 0.05, \quad r_{\rm des} = r_{\rm sdss} + 0.088, \\ i_{\rm des} &= i_{\rm sdss} + 0.112, \quad z_{\rm des} = z_{\rm sdss} + 0.159, \end{aligned} \qquad \text{(B1)}$$

for the `psf` magnitudes, and

$$\begin{aligned} g_{\rm des} &= g_{\rm sdss} + 0.165 - 0.092(g_{\rm des} - r_{\rm des} - 0.4) \\ r_{\rm des} &= r_{\rm sdss} + 0.118 - 0.215(g_{\rm des} - r_{\rm des} - 0.4) \\ i_{\rm des} &= i_{\rm sdss} + 0.04 - 0.2(i_{\rm des} - z_{\rm des}) \\ z_{\rm des} &= z_{\rm sdss} + 0.078 - 0.044(z_{\rm des} - Y_{\rm des} - 0.17) \end{aligned} \qquad \text{(B2)}$$

for the `model` magnitudes. There is considerable scatter $(0.11 - 0.18\mathrm{mag})$ in the translated magnitudes, given by the extendedness of the objects and different depth and image quality between SDSS and DES. This means that interesting candidates (resp. contaminants) can leak out of (resp. within) the hyperplanes defining class boundaries as identified by the ANN classification.

AFFILIATIONS

[1]European Southern Observatory, Karl-Schwarzschild-Strasse 2, 85748 Garching bei München, DE
[2]Fermi National Accelerator Laboratory, Batavia, IL 60510
[3]Department of Physics and Astronomy, PAB, 430 Portola Plaza, Box 951547, Los Angeles, CA 90095-1547, USA
[4]Laboratoire d'Astrophysique, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, CH-1290 Versoix, Switzerland
[5]Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK
[6]Institute for Astronomy, Department of Physics, ETH Zurich, Wolfgang-Pauli-Strasse 27, 8093, Zurich, Switzerland
[7]Institute of Astronomy and Astrophysics, Academia Sinica, P.O. Box 23-141, Taipei 10617, Taiwan
[8]Max-Planck-Institut fr Astrophysik, Karl-Schwarzschild-Str. 1, D-85741 Garching, Germany
[9]Institute of Cosmology & Gravitation, University of Portsmouth, Portsmouth, PO1 3FX, UK
[10]Kavli IPMU (WPI), UTIAS, The University of Tokyo,

---

[7] We refer to Agnello et al. (2015b) for the definition of 'blue and extended' in this case.

Kashiwa, Chiba 277-8583, Japan
[11]Department of Physics, University of California Davis, 1 Shields Avenue, Davis, CA 95616, USA
[12]Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 452 Lomita Mall, Stanford, CA 94305, USA
[13]Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
[14]Institut de Ciències de l'Espai, IEEC-CSIC, Campus UAB, Carrer de Can Magrans, s/n, 08193 Bellaterra, Barcelona, Spain
[15]Department of Physics, University of California, Santa Barbara, CA 93106, USA
[16]Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA
[17]ICRA, Centro Brasileiro de Pesquisas Fsicas, Rua Dr. Xavier Sigaud 150, CEP 22290-180, Rio de Janeiro, RJ, Brazil
[18]Staples High School, Westport CT
[19]CAPES Foundation, Ministry of Education of Brazil, Brasília - DF 70040-020, Brazil
[20]MIT Kavli Institute for Astrophysics and Space Research, 37-664G, 77 Massachusetts Avenue, Cambridge, MA 02139
[21]Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA
[22]Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, Casilla 603, La Serena, Chile
[23]Department of Physics & Astronomy, University College London, Gower Street, London, WC1E 6BT, UK
[24]Department of Physics and Electronics, Rhodes University, PO Box 94, Grahamstown, 6140, South Africa
[25]CNRS, UMR 7095, Institut d'Astrophysique de Paris, F-75014, Paris, France
[26]Sorbonne Universités, UPMC Univ Paris 06, UMR 7095, Institut d'Astrophysique de Paris, F-75014, Paris, France
[27]SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA
[28]Laboratório Interinstitucional de e-Astronomia - LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil
[29]Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil
[30]Department of Astronomy, University of Illinois, 1002 W. Green Street, Urbana, IL 61801, USA
[31]National Center for Supercomputing Applications, 1205 West Clark St., Urbana, IL 61801, USA
[32]Institut de Física d'Altes Energies, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Barcelona, Spain
[33]Institut de Ciències de l'Espai, IEEC-CSIC, Campus UAB, Carrer de Can Magrans, s/n, 08193 Bellaterra, Barcelona, Spain
[34]Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA
[35]Department of Physics, IIT Hyderabad, Kandi, Telangana 502285, India
[36]Excellence Cluster Universe, Boltzmannstr. 2, 85748 Garching, Germany
[37]Faculty of Physics, Ludwig-Maximilians University, Scheinerstr. 1, 81679 Munich, Germany
[38]Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109, USA
[39]Instituto de Fisica Teorica UAM/CSIC, Universidad Autonoma de Madrid, 28049 Madrid, Spain
[40]Department of Astronomy, University of California, Berkeley, 501 Campbell Hall, Berkeley, CA 94720, USA
[41]Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
[42]Department of Physics, The Ohio State University, Columbus, OH 43210, USA
[43]Astronomy Department, University of Washington, Box 351580, Seattle, WA 98195, USA
[44]Australian Astronomical Observatory, North Ryde, NSW 2113, Australia
[45]George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, and Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA
[46]Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, CP 66318, CEP 05314-970, São Paulo, SP, Brazil
[47]Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA
[48]Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain
[49]Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona) Spain
[50]Department of Physics and Astronomy, Pevensey Building, University of Sussex, Brighton, BN1 9QH, UK
[51]Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain
[52]Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA
[53]School of Physics and Astronomy, University of Southampton, Southampton, SO17 1BJ, UK
[54]"Universidade Federal do ABC, Centro de Ciências Naturais e Humanas, Av. dos Estados, 5001, Santo André, SP, Brazil, 09210-580
[55]"Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831

The Astrophysical Journal, 800:11 (23pp), 2015 February 10 doi:10.1088/0004-637X/800/1/11

STRONG LENS TIME DELAY CHALLENGE. II. RESULTS OF TDC1

Kai Liao[1,2,20], Tommaso Treu[2,20], Phil Marshall[3], Christopher D. Fassnacht[4], Nick Rumbaugh[4], Gregory Dobler[5,21],
Amir Aghamousa[6], Vivien Bonvin[7], Frederic Courbin[7], Alireza Hojjati[8,9], Neal Jackson[10], Vinay Kashyap[11],
S. Rathna Kumar[12], Eric Linder[13,14], Kaisey Mandel[11], Xiao-Li Meng[15], Georges Meylan[7], Leonidas A. Moustakas[16],
Tushar P. Prabhu[12], Andrew Romero-Wolf[16], Arman Shafieloo[6,17], Aneta Siemiginowska[11],
Chelliah S. Stalin[12], Hyungsuk Tak[15], Malte Tewes[18], and David van Dyk[19]
[1] Department of Astronomy, Beijing Normal University, Beijing 100875, China
[2] Department of Physics, University of California, Santa Barbara, CA 93106, USA
[3] Kavli Institute for Particle Astrophysics and Cosmology, P.O. Box 20450, MS29, Stanford, CA 94309, USA
[4] Department of Physics, University of California, 1 Shields Avenue, Davis, CA 95616, USA
[5] Kavli Institute for Theoretical Physics, University of California Santa Barbara, Santa Barbara, CA 93106, USA
[6] Asia Pacific Center for Theoretical Physics, Pohang, Gyeongbuk 790-784, Korea
[7] EPFL, Lausanne, Switzerland
[8] Department of Physics and Astronomy, University of British Columbia,
6224 Agricultural Road, Vancouver, BC V6T 1Z1, Canada
[9] Department of Physics, Simon Fraser University, 8888 University Drive, Burnaby BC V5A1S6, Canada
[10] University of Manchester, School of Physics & Astronomy, Jodrell Bank Centre for Astrophysics, Manchester M13 9PL, UK
[11] Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA
[12] Indian Institute of Astrophysics, II Block, Koramangala, Bangalore 560034, India
[13] Lawrence Berkeley National Laboratory and University of California, Berkeley, CA 94720, USA
[14] Korea Astronomy and Space Science Institute, Daejeon 305-248, Korea
[15] Department of Statistics, Harvard University, 1 Oxford Street, Cambridge, MA 02138, USA
[16] Jet Propulsion Laboratory, California Institute of Technology, M/S 169-506,
4800 Oak Grove Drive, Pasadena, CA 91109, USA
[17] Department of Physics, POSTECH, Pohang, Gyeongbuk 790-784, Korea
[18] Argelander-Institut für Astronomie, Auf dem Hügel 71, D-53121 Bonn, Germany
[19] Department of Mathematics, Imperial College London, London SW7 2AZ, UK
Received 2014 August 27; accepted 2014 December 8; published 2015 February 3

ABSTRACT

We present the results of the first strong lens time delay challenge. The motivation, experimental design, and entry level challenge are described in a companion paper. This paper presents the main challenge, TDC1, which consisted of analyzing thousands of simulated light curves blindly. The observational properties of the light curves cover the range in quality obtained for current targeted efforts (e.g., COSMOGRAIL) and expected from future synoptic surveys (e.g., LSST), and include simulated systematic errors. Seven teams participated in TDC1, submitting results from 78 different method variants. After describing each method, we compute and analyze basic statistics measuring accuracy (or bias) A, goodness of fit χ^2, precision P, and success rate f. For some methods we identify outliers as an important issue. Other methods show that outliers can be controlled via visual inspection or conservative quality control. Several methods are competitive, i.e., give $|A| < 0.03$, $P < 0.03$, and $\chi^2 < 1.5$, with some of the methods already reaching sub-percent accuracy. The fraction of light curves yielding a time delay measurement is typically in the range $f = 20\%–40\%$. It depends strongly on the quality of the data: COSMOGRAIL-quality cadence and light curve lengths yield significantly higher f than does sparser sampling. Taking the results of TDC1 at face value, we estimate that LSST should provide around 400 robust time-delay measurements, each with $P < 0.03$ and $|A| < 0.01$, comparable to current lens modeling uncertainties. In terms of observing strategies, we find that A and f depend mostly on season length, while P depends mostly on cadence and campaign duration.

Key words: gravitational lensing: strong – methods: data analysis

1. INTRODUCTION

The past decade has seen the emergence of a concordance cosmology, ΛCDM, in which the contents of the universe are dominated by dark matter and dark energy. Even though the basic parameters appear to be robustly measured, more stringent measurements are sought as a way to improve our understanding of the nature of these mysterious components, as well as a way to test the model against signatures of new physics (Suyu et al. 2012; Weinberg et al. 2013).

Achieving better cosmography means two things. On the one hand, increasingly higher quality data are being obtained (e.g., Planck Collaboration et al. 2014) in order to improve the precision of each method. On the other hand, independent observational methods are being exploited to break the degeneracies inherent to each method and to uncover unknown systematic uncertainties, thus improving accuracy. With precision and accuracy rigorously under control, potential inconsistencies might reveal new physics, such as the presence of additional families of neutrinos or deviations from general relativity.

In the past few years, strong lens time delays (Refsdal 1964; Kochanek 2002) have made something of a comeback, becoming an increasingly popular probe of cosmography (Oguri 2007; Coe & Moustakas 2009; Dobke et al. 2009; Paraficz & Hjorth 2010; Treu et al. 2013; Sereno & Paraficz 2014).

---

[20] Also at Department of Physics and Astronomy, University of California, Los Angeles CA 90095, USA.

[21] Also at Center for Urban Science + Progress, New York University, Brooklyn, NY 11201, USA.

The configuration most suitable for this work consists of a quasar with variable luminosity, being lensed by a foreground elliptical galaxy that creates multiple images of the quasar (e.g., Treu 2010, for a recent review). Differences in optical paths and gravitational potentials give rise to time delays between the images. In turn, the observable time delays, combined with a model of the mass distribution in the main deflector and along the line of sight, provide information on the so-called time-delay distance, which is a combination of angular diameter distances. The time delay distance is primarily sensitive to the Hubble constant (Suyu et al. 2013), but can also constrain other cosmological parameters, especially with large numbers of time delay systems and in combination with other methods (Paraficz & Hjorth 2009; Linder 2011).

At the time of writing, only a fraction of the hundred or so known gravitationally lensed quasars has well-measured time delays, owing to the considerable observational challenge associated with this measurement. Accurate time delays in the optical require long and well-sampled light curves as well as sophisticated algorithms that account for data irregularities and astrophysical effects such as microlensing (e.g., Tewes et al. 2013a). Radio wavelength light curves have been used to determine time delays with great accuracy (e.g., Fassnacht et al. 2002), but unfortunately are restricted to the radio-loud subset of systems. In all cases, the success rate is limited by the intrinsic variability of the sources.

The number of systems with known time delays is about to increase dramatically. In the immediate future, as more lensed quasars are discovered (e.g., via the STRIDES program[22]), there will be more opportunities to identify highly variable systems in cosmologically favorable configurations for targeted follow-up. The state-of-the-art project COSMOGRAIL[23] with its newly developed methods (Tewes et al. 2013a) has shown the potential power of extracting time delay data from sparsely sampled photometric data (Tewes et al. 2013b). In the near future, the upcoming cadenced optical imaging surveys will provide light curves for large samples of lensed quasars. For example, the Large Synoptic Survey Telescope (LSST; LSST Science Collaboration et al. 2009; Ivezic et al. 2008) will repeatedly observe approximately 18,000 deg^2 of sky for 10 years, and is predicted to find and monitor several thousand time delay lens systems (Oguri & Marshall 2010; LSST Dark Energy Science Collaboration 2012).

In preparation for this wealth of light curves, it is crucial to carry out a systematic study of the current algorithms for time delay determination. Such an investigation has two main goals. The first is to determine whether current methods have sufficient precision and accuracy to exploit the kind of data anticipated in the next decade. Identifying limitations and failure modes of current methods is a necessary step to develop the next generation of measurement algorithms. In parallel, the second goal is to test the impact of different observational strategies. For example, what kind of cadence, duration, and sensitivity is required to obtain precise and accurate time delays? Is the LSST baseline strategy sufficient to meet the goals of time delay cosmography or can we identify changes that would improve the outcome?

With these two goals in mind, a time delay challenge (TDC) was initiated in 2013 October. The challenge "Evil" Team (G.D., C.D.F., K.L., P.J.M., N.R., T.T.) simulated large numbers of time delay light curves, including all anticipated physical and experimental effects. The wider community was then invited to extract time delay signals from these mock light curves, blindly, using their own algorithms as "Good Teams."[24] This invitation was made by the posting of an initial version of Paper I of this series (Dobler et al. 2014) on the arxiv.org preprint server, and on the TDC Web site (http://timedelaychallenge.org/).

The two first ladders of this challenge are TDC0 and TDC1. TDC0 consisted of a small set of simulated data, which was used mostly as a debugging and validation tool. TDC0 is discussed in detail in Paper I. Four statistics were used to evaluate the performance of every method's submitted time delays $\tilde{\Delta t}_i$ and uncertainties δ_i, in light of the true time delay value (defined as positive in the input), Δt_i. These four metrics are as follow: the success fraction

$$f \equiv \frac{N_{\mathrm{submitted}}}{N}, \tag{1}$$

where N is the total number of light curves available for analysis in the ladder; the χ^2 value:

$$\chi^2 = \frac{1}{fN} \sum_i \left(\frac{\tilde{\Delta t}_i - \Delta t_i}{\delta_i} \right)^2 ; \tag{2}$$

the "precision"

$$P = \frac{1}{fN} \sum_i \left(\frac{\delta_i}{\Delta t_i} \right); \tag{3}$$

and the "accuracy" or "bias"

$$A = \frac{1}{fN} \sum_i \frac{\tilde{\Delta t}_i - \Delta t_i}{\Delta t_i}. \tag{4}$$

In addition to the sample metrics we also define the analogous metrics for each individual point A_i, P_i, and χ_i^2. Thus, A, P, and χ^2 defined above are the averages of the individual point values.

Target thresholds in each of these sample metrics were set for the teams entering TDC0. The seven "Good" Teams whose methods passed these thresholds were given access to the TDC1 data set, which consisted of several thousand light curves. This large number was motivated by the goals of revealing the potential biases of each algorithm at the sub-percent level and testing the ability of current pipelines to handle large volumes of data.

To put this challenge in a cosmological context, absolute distance measurements with 1% precision and accuracy are highly desirable for the study of dark energy (Suyu et al. 2012; Weinberg et al. 2013) and other cosmological parameters. Therefore, in order for the time delay method to be competitive it has to be demonstrated that the delays can be measured with sub-percent accuracy *and* that the combination of precision for each system and the available sample size is sufficient to bring the statistical uncertainties to sub-percent level in the near future. The total uncertainty on the time delay distance, and therefore on the derived cosmology, depends on both the time delay and on the residual uncertainties from modeling the lens potential and the structure along the line of sight. Thus, controlling the precision and accuracy of the time delay measurement is a

---

[22] strides.physics.ucsb.edu

[23] http://www.cosmograil.org

[24] We note here that the tongue-in-cheek names "evil" and "good" teams do not denote any despicable intention or moral judgment, but were chosen to capture the desire of the challenge designers to produce significantly realistic (and difficult) light curves as well as an incentive for the outside teams to participate.

Table 1
The Observing Parameters for the Five Rungs of TDC1

Rung	Mean Cadence (days)	Cadence Dispersion (days)	Season (months)	Campaign (yr)	Length (epochs)
0	3.0	1.0	8.0	5	400
1	3.0	1.0	4.0	10	400
2	3.0	0.0	4.0	5	200
3	3.0	1.0	4.0	5	200
4	6.0	1.0	4.0	10	200

necessary, but not sufficient, condition. In this first challenge we focus on just the time delay aspect of the measurement. The assessment of residual systematic uncertainties in the other components of time delay lens cosmography, and the distillation of the time delay measurement biases and uncertainties into a single cosmology metric is left for future work.

This paper focuses on TDC1, the analysis period of which closed on 2014 July 1, and it is structured as follows. Section 2 contains a brief recap of the light curve generation process, and describes the design of TDC1. In Section 3 we describe the response of the community to the challenge and give a brief summary of each method that was applied, and then in Section 4 we analyze the submissions. We look at some of the apparent implications of the TDC1 results for future survey strategies in Section 5, and briefly discuss our findings in Section 6. In Section 7 we summarize our conclusions.

2. DESCRIPTION OF TIME DELAY CHALLENGE TDC1

In TDC1, the "Evil" Team simulated several thousand realistic mock light curve pairs, using the methods outlined in Paper I. In this section, we first describe the general five rung design of TDC1, and then describe the process of generating these light curves step by step, revealing quantitative details of all the elements considered. We emphasize that TDC1 was purely a light curve analysis challenge; no additional information regarding the gravitational lensing configuration, such as positions of the multiple images, or redshifts of the source and deflector, was given. This choice was motivated by the goal of performing the simplest possible test of time delay algorithms. As discussed at the end of this paper, the inclusion of additional lensing information could provide means to further improve the performance of the methods.

2.1. The Rungs of the Challenge

Each rung of TDC1 represents a possible wide-field survey that has monitored sufficient sky area that we are in possession of light curves for 1000 gravitationally lensed active galactic nucleus (AGN) image pairs. The number of lens systems in this sample is somewhat less than 1000: quad systems are presented as two pairs, flagged as coming from the same system but enabling two independent time delay measurements. The five rungs of TDC1 span a selection of possible observing strategies, ranging from a high cadence, long season dedicated survey (such as COSMOGRAIL might evolve into), to the kind of "universal cadence" strategy that might be adopted for an "all-sky" synoptic imaging survey (such as is being designed for LSST). The challenge allows four control variables to be investigated (within small plausible ranges): cadence, sampling regularity, observing season length, and campaign duration. Table 1 gives the values of these control variables for each rung.

To make the mock data generation more efficient, and to better enable comparison of results between the different rungs, we re-used the same catalog of lenses for all the rungs. This trick was disguised from the "Good" Teams by randomly re-allocating the light curve identification labels in each rung. In addition, the random noise was independently generated in each rung. As a consequence, the submissions for different rungs may be deemed independent, as if they had addressed 5000 lensed image pairs.

2.2. Lens Sample

The time delays between the light curves of gravitationally lensed images are determined primarily by the macro structure of the lens galaxy. For the TDC1 sources and lenses we use the mock LSST catalog of lensed quasar systems prepared by Oguri & Marshall (2010, hereafter OM10).[25] This sample was drawn from plausible physical distributions for the various key properties of lensed quasar systems and very approximate observing conditions expected with LSST, namely, a characteristic angular resolution of 0.75 arcsec and a 10σ limiting magnitude per monitoring epoch of 23.3 in the i band. Assuming a survey area of 18,000 deg^2, these numbers correspond to an OM10-predicted mock sample of some 2813 lenses. Given these constraints, we randomly drew 720 doubly imaged and 152 quadruply imaged quasars from this catalog, to give a total of 1024 independent time delayed image pairs. As Figure 1 shows, the mean time delay in TDC1 is several tens of days. We rejected all time delays outside the range 5 to 120 days as we drew the mock sample, since the typical observing cadence and season length are expected to be a few days and a few months, respectively. The same time delay range constraint reduced the parent OM10 mock lens sample by 76%, to 2124 lenses. When analyzing the submissions, we found that very few accurate measurements of time delays less than 10 days were possible, and so in the rest of this paper we focus on the range $10 < \Delta t < 120$ days. Imposing this narrower range on the OM10 mock LSST lens sample results in 1990 systems. While the image pairs with 5 days $< \Delta t <$ 10 days were not used in the analysis, they are still there in the TDC1 data set for potential future use.

To give an overview of this sample, we show the distributions of time delays Δt between images in our 1024 image pairs (in Figure 1), and detection magnitudes i_3 in the 872 lens systems (in Figure 2). The i_3 quantity is the i-band magnitude of the third brightest image in a quad system or the magnitude of the fainter image in a double-image system. (It is an important parameter because it helped OM10 characterize the detectability of lensed quasars: lenses are assumed to be measurable if i_3 is above the 10σ limiting magnitude of a survey.) The lens abundance rises fairly steeply with i_3, so in order to probe the relationship

---

[25] The OM10 catalog is available from https://github.com/drphilmarshall/OM10.

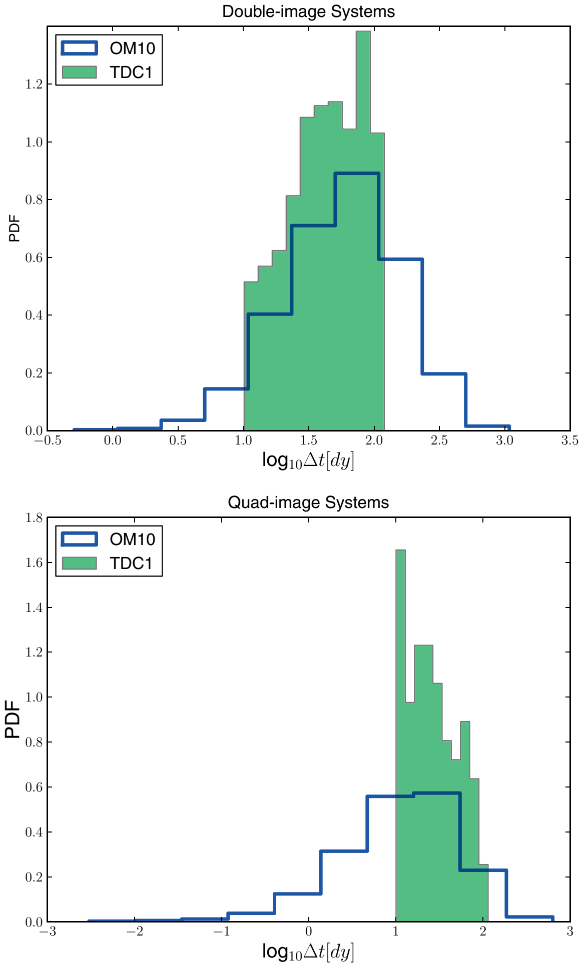


Figure 1. Time delay distributions, from both the parent OM10 catalog and the sample used in the TDC1 analysis, for the double-image (top) and quad-image (bottom) systems.

between it and the time delay measurement accuracy, we split the magnitude range 20–24 into four sub-ranges, and selected approximately equal numbers of systems in each sub-range.

In summary, our sample is similar to OM10's, except that the brighter lenses and intermediate time delays are somewhat over-represented. As we will discuss later in this paper, this allows us to sample the range of magnitudes more evenly, while introducing negligible bias in the inferred performance of the methods.

2.3. Generation of Intrinsic Light Curves

The mechanism for generating intrinsic light curves is described in Paper I. In TDC1, we needed to simulate many more data sets; the most time-consuming part was generating the damped random walk (DRW) stochastic process with which we modeled the intrinsic AGN light curves. The interval between discrete epochs had to be 0.01 days in order to enable the counter-image light curve to be simulated with a time delay precision sufficient to not affect the ensemble metrics. Each of these intrinsic light curves took approximately 1–2 CPU hours to make, so for efficiency we created just 500 intrinsic light curves, each of 10 yr length, and re-cycled them between several mock data sets, with different starting epochs chosen relative to the season gaps, so that all the release data could be considered to be independent.

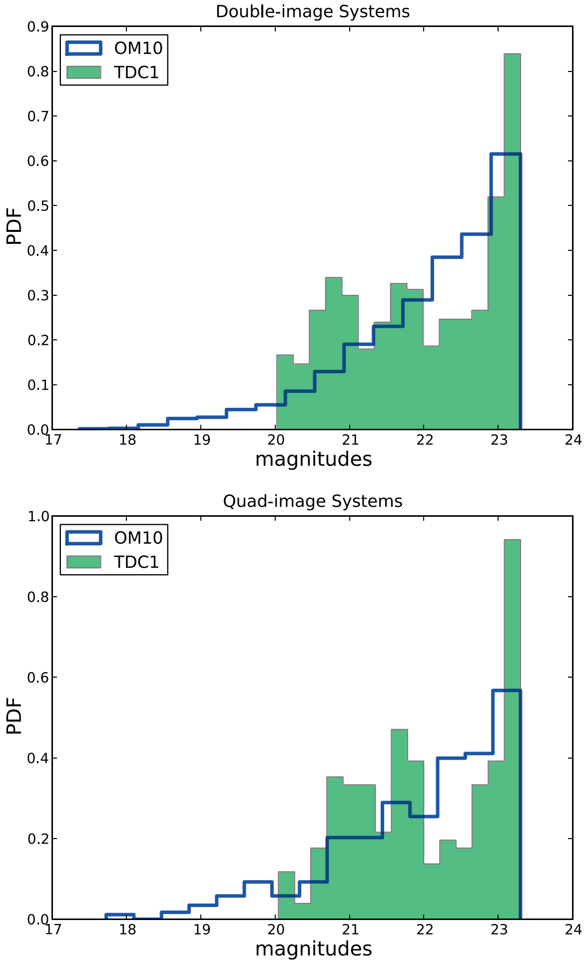


Figure 2. Detection magnitude "i_3" distributions for the double (top) and quad (bottom) systems. For doubles, i_3 is the magnitude of the fainter image, while for quad systems it is the magnitude of the third-brightest image. Distributions are shown both for the parent OM10 sample, and the sample used for TDC1.

The DRW light curves represent light curve fluctuations, and have zero mean magnitude. They are determined by only two parameters: the characteristic timescale τ and the characteristic amplitude of the fluctuations σ. These were drawn from distributions designed to match that observed for the spectroscopically confirmed ($i < 19.1$ mag) quasars in MacLeod et al. (2010). Their $\log \tau$ and $\log \mathrm{SF}_\infty$ (asymptotic rms variability on long timescales) parameters were drawn uniformly from the ranges [1.5 : 3.0] and [−1.1: − 0.3], respectively. The endpoints of these ranges correspond to 30 and 1000 days, and 0.08 and 0.5 mag. The rms fluctuation level was derived for each light curve via $\sigma = \mathrm{SF}_\infty/\sqrt{\tau}$.

2.4. Modeling Microlensing

Microlensing is an important source of systematic error because it makes the multiply imaged light curves differ by more than the time delay and the macrolens magnification ratio. In galaxy-scale lenses, the variability of the microlensing typically has a timescale significantly larger than that of the quasar intrinsic variability (although occasional caustic crossing events can provide some transient rapid variability). We expect the most successful light curve measurement algorithms to model an additional microlensing light curve component individually at each image.

Given an OM10 catalog convergence κ, shear γ and surface density in stars F_* at each image position, we generated a static stellar field with a mean mass per star of $0.3\,M_\odot$ (Schechter et al. 2004). We then calculated its source plane magnification map and convolved this with a Gaussian kernel to represent the extended accretion disk of the source quasar; we drew source sizes s (Gaussian radii) uniformly from the range $[10^{14}\text{–}10^{16}]$ cm. When calculating the microlensing light curves, we assumed Gaussian distributions for the components of the relative velocity v between the source and the stars in the lens, with standard deviation of $500\,\mathrm{km\,s^{-1}}$ in each direction.[26] In the Appendix we show how the scatter in microlensing variability amplitude depends on F_*, κ, and source size. Finally, we note that there are several characteristic timescales in microlensing light curves, ranging from the crossing time of the mean stellar mass Einstein Radius (Paraficz et al. 2006) to the source caustic crossing time, to the density of caustics in the network, and those can give rise occasionally to quasi-periodic features.

2.5. Photometric and Systematic Errors

Following Tewes et al. (2013a) we considered several sources of observational error when generating the light curve fluxes. The main source of statistical uncertainty is the sky brightness, which we assume dominates the photometry. We used the approximate distribution of 5σ limiting point source magnitudes from one of the LSST project operations simulator outputs (L. Jones 2014, private communication), and converted these to flux uncertainties. The mean and standard deviation of the 5σ i-band limiting flux was found to be 0.263 and 0.081 AB nanomaggies[27], respectively; to add photometric noise to a light curve flux we first drew an rms photometric uncertainty from a Gaussian of mean 0.053 and width 0.016 nanomaggies (dividing the above numbers by 5), and then drew a noise value from a Gaussian of width equal to this rms. The minimum noise value was set to be 0.001 nanomaggies.

Beyond this basic (though possibly epoch-dependent) Gaussian noise, we might expect additional flux errors to be present as the observing set-up changes over a long monitoring campaign. To mimic such fluctuations, we added the following three types of "evilness" to the light curves.

1. Flux uncertainty under-estimation: for each pair of light curves and for approximately 1 in every 10 epochs, we added noise that was three times larger than standard, but reported it as the normal one.
2. Calibration error: for each pair of light curves and for approximately 1 every 10 epochs, we added correlated noise, i.e., both points were higher or lower than in the normal case.
3. Episodic transparency loss: we took a subset of the data (a few weeks every year), and offset the fluxes by 1% or 3%.

There could be more than one type of "evilness" present in any given light curve: the combinations applied to the TDC1 light curves were as follows. 3% of the light curves, selected randomly, were contaminated with a single type of "evilness." Another 1% were contaminated with two types, and 3% were contaminated with all three. In total then, 15% of the light curves were contaminated with these simulated bad observational conditions.

2.6. Example TDC1 Light Curves

Figures 3 and 4 illustrate the process of generating TDC1 data in each of the five rungs, using light curves selected randomly from those data sets. The top panels show the AGN intrinsic light curves in magnitudes. The panels beneath them show the microlensing magnifications (also in magnitudes). The third panels show the AGN light curves with microlensing effects, and the effect of sampling is shown in the fourth panels. Finally, the sparsely sampled noisy mock light curves are shown on the bottom panels, in flux units.

Comparing panels 3 and 5, we can easily see how two similar curves become difficult to associate by eye once the sparse sampling and the addition of noise have been applied. Table 2 shows the values of the input parameters τ, σ, v, s, F_*, enabling some intuition to be developed by comparing plots shown for the different rungs.

3. RESPONSE TO THE CHALLENGE

As described in Section 1, the TDC was presented to the community as two "ladders," TDC0 and TDC1. The TDC0 data were used as a gateway to TDC1; in order to gain access to the TDC1 data, each "Good" Team had to submit a set of time delays inferred from TDC0 that met the targets described in Section 1, and in more detail in Paper I. In total, 13 "Good" Teams participated in TDC0, many of which submitted multiple sets of solutions. Seven teams passed TDC0 and, went on to participate in TDC1. One of the teams submitted results based on three different algorithms: those were considered independent submissions. In addition, the "Evil" Team did an in-house analysis of the TDC1 data, using a relatively simple procedure, to serve as a baseline comparison for the "Good" Team submissions. All 10 of these algorithms are described below and some of their properties are summarized in Table 3. It is worth noting that the teams continued to develop their methods between TDC0 and TDC1 and beyond, and the description given here is for the versions of the methods that were applied to TDC1.

3.1. Benchmark Technique by Rumbaugh ("Evil" Team)

The baseline method used by the "Evil" Team was a χ^2-based Markov Chain Monte Carlo (MCMC) approach. While the member of the team that wrote and executed this baseline method (N.R.) did not work directly on simulating the light curves, this method should not be considered blind in the same way as the "Good" Teams'.

In practice the method consists of comparing a shifted copy of one of the light curves to the other light curve, and using a χ^2 function to compute the posterior probability distribution function for the time delay. Matching the light curves requires some interpolation, which was carried out using a boxcar kernel

---

[26] The microlensing code used in this work, MULES, is freely available at https://github.com/gdobler/mules.

[27] One "AB maggy" is the flux corresponding to an AB magnitude of 0.0 (Stoughton et al. 2002). Thus, 0.263 nanomaggies is the flux corresponding to an AB magnitude of 24.

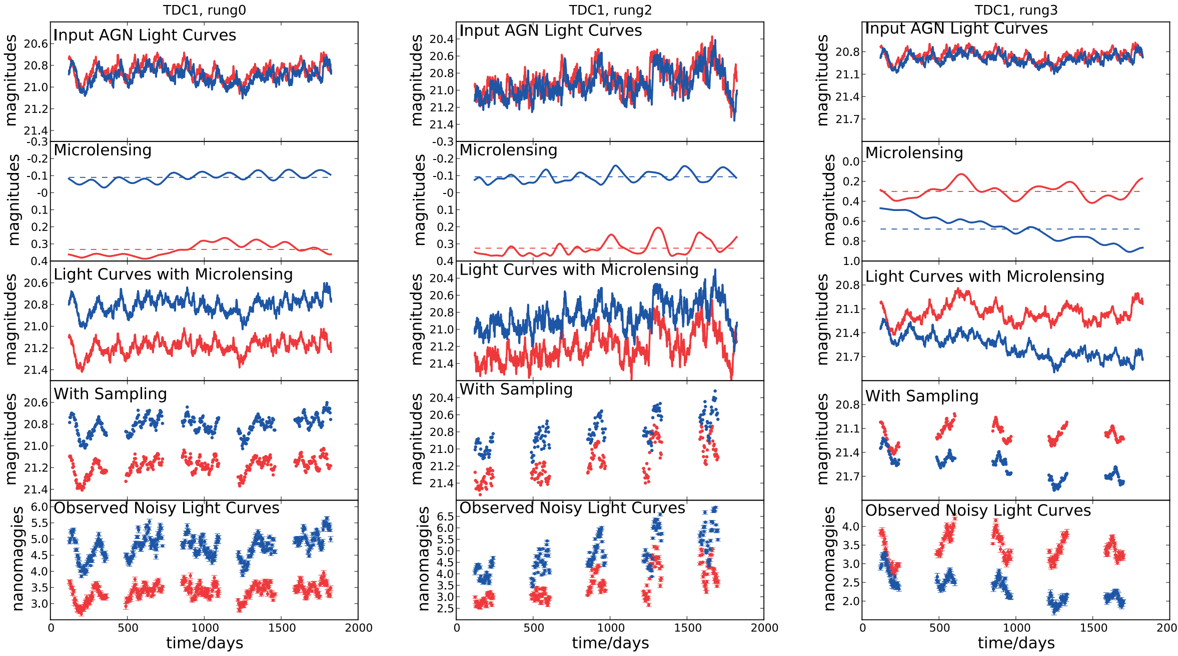


Figure 3. Illustration of the process of generating time delay light curves, with examples taken from the Rung 0 (left), Rung 2 (middle), and Rung 3 (right) samples. The panels in each figure show, going from the top to the bottom, (1) the input AGN light curves, (2) the microlensing contributions in magnitudes, (3) the AGN light curves including the microlensing contributions, (4) the result of down-sampling to the required cadence and season length, and (5) the final sparsely sampled noisy light curves.

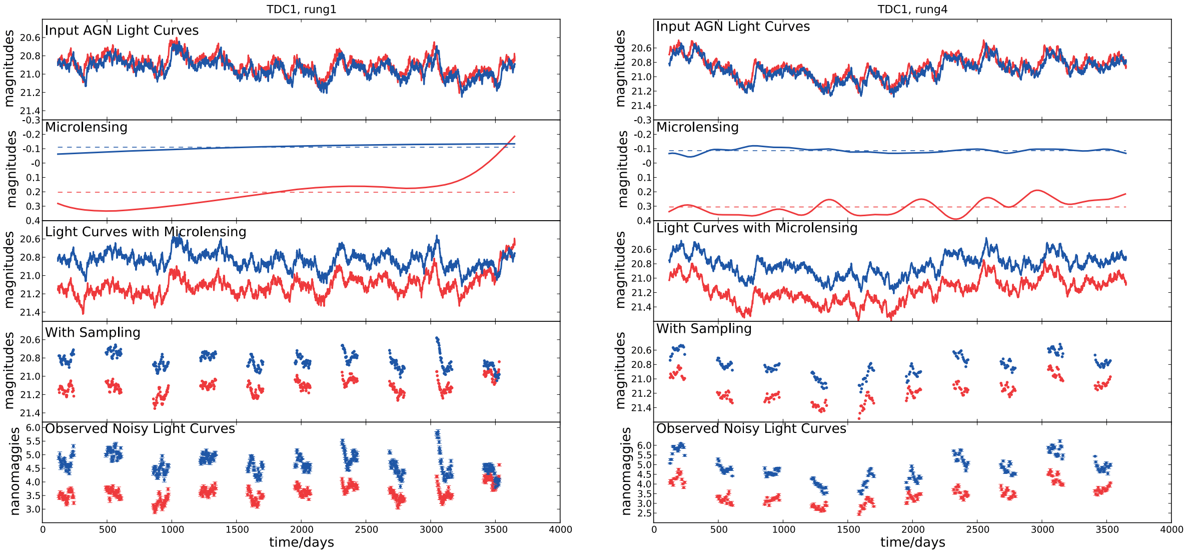


Figure 4. Same as Figure 3, but for the longer campaign-duration light curves of Rungs 1 and 4.

Table 2
The Parameters Used to Make the Simulated Data Shown in Figures 3 and 4, to Enable Study of Their Effects on the Light Curves

Rung	τ (days)	σ (mag/day$^{-1/2}$)	v (km s^{-1})	s (10^{14} cm)	F_{*A}	F_{*B}
0	37.8	0.017	731	3.87	0.037	0.062
1	83.0	0.017	731	38.7	0.037	0.062
2	40.6	0.039	1462	3.87	0.037	0.062
3	37.8	0.017	731	3.87	0.019	0.031
4	178.0	0.017	365	3.87	0.037	0.062

Table 3
Summary of Methods Explicitly Accounting for Microlensing

Method	Microlensing
Rumbaugh	No
Shafieloo	Yes
PyCS-d3cs	Yes
PyCS-sdi	Yes
PyCS-spl	Yes
Jackson-manchester	Yes
Kumar	Yes
JPL	No
Hojjati	Yes
DeltaTBayes	No

with a full width of 10 days. This particular kernel was chosen to save computational time; however, the choice of the kernel did not have a significant effect on the accuracy or precision of the method. In order to gain additional computational speed, the correlation between temporally close data points introduced by the smoothing kernel was neglected. This approximation reduced the computation time by about an order of magnitude, while providing only marginally worse accuracy. The posterior was sampled using the *emcee* (Foreman-Mackey et al. 2013) software package. For each trial value of the time delay, only the overlapping parts of the time-shifted light curves were used in the computation of the change in χ^2. To avoid calculations using small overlap regions, a maximum time delay was imposed equal to 75% of the shortest season length of the data set currently being analyzed. Time delay point estimates were chosen to be the median of the output sample values, with the uncertainties chosen to be half the width of the region containing 68.3% of the chain surrounding the median.

Before applying the benchmark technique to TDC1 data, it was tested on the TDC0 data, as well as on an additional set of simulated data designed to be similar to TDC0. In this testing, the smoothing kernel was varied, as well as several other aspects of the method as indicated above (including whether or not the full covariance matrix was used). The accuracy and precision of the inference were found to not depend significantly on these choices.

Time delay estimates from three implementations of this method were submitted, with the aim of producing answers of different degrees of reliability. The three implementations were obtained by restricting the submissions to those systems with estimated time delay uncertainty below 6, 10, and 20 days. The submissions resulting from these cuts are named Gold, Silver, and Bronze, respectively.

3.2. Gaussian Processes by Hojjati & Linder

This "Good" Team implemented Gaussian process (GP) regression to estimate the time delays (see Hojjati et al. 2013, for the basic approach). GPs are widely used as a model-independent technique for reconstructing an underlying function from noisy measurements. The GP is specified by a mean function, and a covariance (kernel) function characterized by a set of hyperparameters, describing the time delay, relative magnitude shift, QSO variability and coherence length, microlensing variability and coherence length, and measurement noise. This approach is very flexible, not assuming a physical model for the quasar or microlensing input, but allowing the data to decide how best to describe the signal in terms of a GP. The hyperparameters were fitted to data using the GP likelihood through a Bayesian analysis. The parallel and highly efficient fitting code employed two covariance kernels, two optimization methods, and variation of priors to cross-check the results for robustness. The team passed or rejected a system, based on the consistency of fits and their likelihood weights, and then assigned a final best fit, uncertainty, and confidence class to the passed systems.

The overall philosophy emphasized complete automation and accuracy of estimation, rather than precision (e.g., fitting down to five day delays and placing no cut on precision) or numbers of fits. Within this, the team fine-tuned samples based on their confidence in the fit, and to a lesser extent the error estimation. Six samples were submitted, with the basic three representing progressively more inclusive fit confidence along the lines of, e.g., gold, silver, bronze estimation. These correspond to the samples nicknamed Lannister, Targaryen, and Baratheon, respectively. In addition, a more conservative sample (nicknamed Tully) and one with tighter error assignment (nicknamed Stark) were submitted. Catastrophic outliers were identified by running selected samples (e.g., especially short or long time delays) with controlled priors, and also an analysis of the best-fit parameters for the selected systems. The sample nicknamed "Freefolk" was the result of such analysis.

A correction to the mean function treatment in the code significantly increased the consistency of the fits. However, since this modification was made after the TDC1 submission deadline, this is not reflected in the results presented in this paper; see the updates and discussion by Hojjati & Linder (2014). Furthermore, the method has benefited from, and was improved after, a reanalysis of the fits and the investigation of the hyperparameter behavior using the unblinded TDC1 data.

3.3. FOT by Romero-Wolf & Moustakas

The Full of Time (FOT) team's GP inference algorithm took a Bayesian approach to solve for the delay between a pair of light curves. The probability of the light curve parameters $\bar{M}$ (mean magnitude), σ (characteristic amplitude of the fluctuations), and τ (characteristic timescale) given the data is proportional to the product of the likelihood function for a CAR process (Kelly et al. 2009; MacLeod et al. 2010) and uniform priors. Details about the CAR process can be also found in Paper I. The *emcee* (Foreman-Mackey et al. 2013) MCMC ensemble sampler provides an estimate of the posterior probability distribution for the light curve parameters. To reconstruct the delay, the pair of light curves were combined into a single time series assuming a delay and magnitude offset. The probability of the delay and magnitude offset, along with light curve parameters, is given by the CAR process likelihood function of the combined light curve and uniform priors. The light curve delay and its uncertainty were then inferred from the marginalized posterior distribution for the time delay given the light curves. The algorithm did not characterize or fit for microlensing, although it identifies the data sets that are most likely to have microlensing variations. A more thorough description of this method and internal tests are being written up by L. A. Moustakas & A. Romero-Wolf (2014, in preparation).

The procedure was tested by generating tens of thousands of "blind" time-delayed light curves through the CAR process, with varying (irregular) observational patterns and campaigns, photometric uncertainties, magnitude offsets, and time delays. These were then processed with the inference technique described above. Both the successful recovery rate and the precision of the (marginalized) time delay and magnitude offset were then studied as a function of each "observational" parameter (i.e., the observational campaign factors and the assumed photometric precision).

To avoid outliers, a set of consistency requirements between the posterior distributions for the individual and combined light curve parameters were required. A solution was rejected if the mean of the posterior σ distributions from each light curve and their combinations differed by more than 2.6 root-sum-squared standard deviations. The means of the posterior $\log_{10}\tau$ distributions for each light curve must also agree to within one standard deviation, forcing a consistency in the physical behavior of the reconstructed "stitched" data set compared to the input data. Additional quality cuts were included from inspection of the reconstructed time delay and time delay uncertainty scatter relation. These required that delays less than 100 days have uncertainties smaller than 10 days. The ratio of the delay uncertainty to the delay was also required to be smaller than 2.

3.4. Smoothing and Cross-correlation by Aghamousa & Shafieloo

This "Good" Team combined various statistical methods of data analysis in order to estimate the time delay between different light curves. At different stages of their analysis they used iterative smoothing, cross-correlation, simulations and error estimation, bias control, and significance testing to prepare their results. Given the limited timeframe (they started the project in early 2014 May), they had to make some approximations in their error analysis.

In their approach to estimate the time delay between a pair of light curves A_1 and A_2, they first smoothed over both light curves using an iterative smoothing method (Shafieloo et al. 2006; Shafieloo 2007, 2012; Shafieloo & Clarkson 2010), producing the smoothed light curves A_1^{smooth} and A_2^{smooth}. During smoothing, they recorded the ranges with no data available (which would have resulted in unreliable smoothing). The algorithm was set to automatically detect such ranges. Then, they calculated the cross-correlation between A_1 and A_2^{smooth} and also between A_2 and A_1^{smooth} for different time delays, and found the maximum correlations. These two maximum correlations should be for the same time delays (that is, the absolute values of the time delays should be consistent with each other). The difference between these two estimated time delays (with maximum correlations) was part of the total uncertainty considered for each pair (in the estimated time delay). To estimate the error on each derived time delay, the team also simulated many realizations of the data for each rung, and for various time delays. Knowing the fiducial values, they derived the expected uncertainties in the estimated values of the time delays.

This team also performed bias control, since long time delays have a limited data overlap between the two light curves. In the case of the quad sample, they used different combinations of the smoothed and raw light curves to test the internal consistency of the results and relative errors. These internal consistency relations can be used to adjust the estimated error-bars for each pair (considering the consistency of all light curves as a prior). The team selected for cross-correlations between the two light curves with more than 50% or 60% correlation coefficients. Pairs with potentially high bias were cut as well. In this methodology the light curves are compared in multi-segments. The effect of micro-lensing can be considered as a linear distortion in these segments. While the correlation coefficient is unchanged under linear transformation, there is no concern for micro-lensing in this algorithm and the method is unaffected. Additional details of this method will be described in a separate paper Aghamousa & Shafieloo (2014).

3.5. Supervised Pelt by Jackson

All pairs of joint light curves were inspected by eye by this team, using a Python tool developed for the purpose. An initial Pelt et al. (1994) statistic was calculated for a large range of time delays, and its minimum found, but this resulted in catastrophic errors in many cases and was frequently over-ridden by visual inspection. Time delays were regarded as believable if (1) at least three coincident points of inflection were detected in the light curves, (2) if no discordant features were seen (i.e., differences between the light curves which could not be plausibly attributed to microlensing) and (3) if the plot of the Pelt statistic against time delay showed a smooth and well-defined minimum.

In the process of assessing the light curves by eye, the following operations were available to find a time delay fitting the above criteria: (1) smoothing of either light curve to match the scatter of the other; (2) adjustment of the zero-point of each segment of the light curve to match the zero-point of the segment of the other light curve that it overlapped using the current time delay; (3) manual adjustment of the current time delay; (4) deletion of one or more segments of the light curve if they were judged to be severely affected by microlensing. In practice, this was the case if a simple rescaling of a whole segment of data between the two light curves produced residuals much larger than those of other rescaled segments. This will happen if the microlensing produces a large change in flux over the period of one data segment; the method therefore roughly corresponds to assuming that microlensing produces variations on a timescale larger than those of the intrinsic brightness variations of the quasar, and deleting regions of data where this is not the case. In most cases, the delay and its error bar were calculated after this process using 100 instances of resampling of the data set using the observed flux errors and a small Gaussian error in each time stamp. This allowed the calculation of a set of delays, in each case using the delay from the Pelt statistic minimum, from which the mean and scatter was used for the delay and its error. In a few cases, mostly those in which the Pelt statistic versus time delay plot had a local minimum around the optimum, an additional error, or in some cases a minor adjustment to the value, was estimated by eye. The error bar was also adjusted in cases where the optimization using smoothing and adjustment of the zero point resulted in a significant reduction of the error estimated by the resampling process. With practice, about 100 pairs of light curves per hour could be processed, so that thousands or tens of thousands of light curve pairs could in principle be analyzed using this method.

The same basic algorithm was used for all submissions, but different submissions were made by separating the objects into three categories, again by eye, according to confidence that the time delay was correct within the stated error. Evaluations with less confidence corresponded to violation of one or more of the believability conditions, and the least certain category usually involved light-curves with only two clearly detected points of inflection. (For each of the three categories, subsidiary submissions were also made with a smaller number of rungs). Three catastrophic errors in rung 0 of the original blind submission were due to incorrect entry of a minus sign during the manual adjustment process in three objects; these were corrected in a non-blind submission which consisted of the original blind submission for all rungs, and all three confidence levels with the three signs corrected. The program was accordingly modified to question the user in the case of large changes imposed by hand.

3.6. *PYCS by Bonvin, Tewes, Courbin, & Meylan*

The PyCS team made submissions using three time delay measurement methods: `d3cs`, `spl`, and `sdi`. The latter two build on initial estimations provided by the former. The following subsections summarize each of these three methods (see V. Bonvin et al., in preparation, for more details).

3.6.1. *d3cs: D3 Curve Shifting*

This first method is based on human inspection of the light curves, in the spirit of citizen science projects. The PyCS team has developed a dedicated browser-based visualization interface, using the D3.js JavaScript library[28] by Bostock et al. (2011). The tool is now publicly available online.[29]

The main motivation behind this time-consuming yet simple approach were to obtain, for each light curve pair, (1) a rough initial estimate for the time delay and its associated uncertainty, and (2) a robust characterization of the confidence that this estimate is not a catastrophic error. The interface asks each user to pick a confidence category for the proposed solution, among four choices:

1. "doubtless" if a catastrophic error can be virtually excluded,
2. "plausible" if the solution yields a good fit and no other solutions are seen,
3. "multimodal" if the proposed solution is only one among two or more possible solutions,
4. "uninformative" if the data does not reveal any delay.

At least two human estimates were obtained for each pair of curves. The database of `d3cs` estimates was then carefully reduced to a single estimate per pair, resolving any conflicts between estimates in a conservative way. A key result of this step was a sample of 1628 "doubtless" time-delay estimates, which the team hoped to be free from any catastrophic outliers. Through this exercise, the team demonstrated that such an approach remains tractable for about 5000 light curves, with typical human inspection times of a minute per light curve pair and user.

3.6.2. *spl: Free-knot Spline Fit*

The `spl` method is a simplified version of the "free-knot spline technique" described by Tewes et al. (2013a) and implemented in the PyCS software package. Using the `d3cs` estimate as the starting point, the method simultaneously fits a single spline representing the intrinsic QSO variability, and a smoother "extrinsic" spline representing the differential microlensing variability, to the light curves. During this iterative process, the curves were shifted in time so as to optimize the fit. The fit was repeated 20 times, starting from different initial conditions, to test and improve the robustness of the resulting delay against local minima of the χ^2 hyper surface. Such a model fit was then used to generate 40 simulated noisy light curves with a range of true time delays around the best-fit solution. By rerunning the spline fit on these simulated curves, and comparing the resulting delays with the true input time delays, the delay measurement uncertainty was estimated.

The `spl` method for TDC1 is simpler, faster, and significantly less conservative in the uncertainty estimation than the free-knot spline technique that was applied to the COSMOGRAIL data[30] by Tewes et al. (2013b) and Rathna Kumar et al. (2013). In particular, the temporal density of spline knots was automatically determined from signal-to-noise ratios measured on the two light curves, and only white noise was used in the generative model. With these simplifications, the team expects the resulting TDC1 error estimates to be rather optimistic. The entire `spl` analysis took about 5 CPU-minutes for an average TDC1 pair.

3.6.3. *sdi*

The third method, `sdi` (for spline difference) was inspired by the "regression difference technique" of Tewes et al. (2013a), replacing the GP regressions by spline fits to speed up the analysis. The method involves fitting a different spline to each of the two light curves, and then minimizing the variability of the difference between these two splines by shifting them in time with respect to each other. The advantage of this approach is that it does not require an explicit microlensing model. To estimate the uncertainty, this method uses the simulated light curves provided by the `spl` technique. As in the `spl` technique, the estimates from `d3cs` were used as the starting point to define the time delay intervals in which `sdi` optimizes its cost function.

3.6.4. *Identification of Catastrophic Failures*

To prevent catastrophic failures, this team relied solely on the `d3cs` "doubtless" sample. The `spl` and `sdi` methods do not alter this confidence classification. Furthermore, a small number of `spl` and `sdi` measurements that did not lie within 1.5σ of the corresponding `d3cs` estimates were rejected.

3.6.5. *Differences between Submissions*

For all three methods, the submissions were named following the scheme `A-B-C-D.dt`, where

A gives the method, `d3cs`, `spl`, or `sdi`;
B gives the method parameters, with `vanilla` denoting the a priori best or simplest;
C gives the confidence category, with `dou` for doubtless and `doupla` for both doubtless and plausible light curve pairs. The `doupla` submissions are expected to be contaminated by some catastrophic outliers, but feature more than twice the number of time delays than the `dou` sample; and
D gives the filter that selects systems according to different criteria across all rungs, mostly based on the blind relative precision $\delta_i/|\widetilde{\Delta t_i}|$. The code `full` corresponds to no filter. `XXXbestP` selects the XXX "best" systems in terms of blind relative precision, `P3percent` selects the largest number of systems so that the average blind relative precision is approximately 3%, and `100largestabstd` is the selection of the 100 largest delays.

Submissions that share the same method and method parameters (A and B) differ only in the selection of systems, and not in the numerical values of the estimates. They can thus be seen as subsamples of the A-B-dou/doupla-full submissions.

---

[28] Data-Driven Documents, http://www.d3js.org/.

[29] https://www.astro.uni-bonn.de/~mtewes/d3cs/tdc1/—see "Read me first" for help.

[30] http://www.cosmograil.org

3.7. *Difference-smoothing by Rathna Kumar, Stalin, & Prabhu*

The difference-smoothing technique, originally introduced by Rathna Kumar et al. (2013), is based on the principle of minimizing the residuals of a high-pass filtered difference light curve between the lensed quasar images. The method is a point estimator that determines an optimal time delay between two given light curves, and an optimal shift in flux to one of the light curves, besides allowing for smooth extrinsic variability. To estimate the uncertainty of the measured time delay in Rathna Kumar et al. (2013), this team made use of simulations produced and adjusted according to Tewes et al. (2013a). However, for participation in the TDC, they made use of a modified version of the difference-smoothing technique as presented by Rathna Kumar et al. (2014). In that paper, they describe an optimal way to adjust the two free parameters in the technique according to the peculiarities of the light curves under analysis and also introduce a recipe for simulating light curves having true delays at discrete intervals in a plausible range around the optimal time delay found. These simulations were used to estimate the uncertainty of the measured value of the time delay. Outliers were identified by noting when the team's technique was found to return random time delays which were uncorrelated with the true delays in their simulated light curves.

The free parameters in the technique are decorrelation length and smoothing timescale. For participation in the TDC, the value of decorrelation length was set equal to the mean temporal sampling of the light curves and the value of smoothing timescale was set equal to the largest integer multiple of decorrelation length for which the amplitude of residual extrinsic variability was less than the 3σ level of noise for each of the light curves. In the absence of significant extrinsic variability between the light curves, the value of smoothing timescale was set equal to ∞.

3.8. Δ *t-Bayes by Tak, Meng, van Dyk, Siemiginowska, Kashyap, & Mandel*

A fully Bayesian approach was developed by this team, based on the key assumption that one of the unobserved underlying light curves is a shifted version of the other. The horizontal shift is the time delay (Δt), and the vertical shift is the magnitude offset (c). Both shifts are treated as unknown parameters. Specifically, from the state-space modeling perspective, it was observed that $\mathbf{x(t)} \equiv \{x(t_1), x(t_2), \ldots, x(t_n)\}$ and $\mathbf{y(t)}$, transformed into the logarithm of flux, around the irregularly sampled underlying light curves, $\mathbf{X(t)}$ and $\mathbf{Y(t)} \equiv \mathbf{X(t} - \Delta t) + c$ each, with measurement errors in log scale. The posterior distribution for Δt is of primary interest. Also, it was assumed that the unobserved true process $\mathbf{X(t)}$ follows an Ornstein-Uhlenbeck process (also known as CAR) as described by Kelly et al. (2009), although a different parameterization was used for more efficient model fitting. Harva & Raychaudhury (2006) proposed a similar idea, but they assumed a different model for the underlying process.

This Bayesian approach treats the unknown parameters as random variables and this team uses specific prior distributions for the time delay and magnitude offset: $p(\Delta t, c) \propto \delta_{\{|\Delta t| \in [0, (t_n - t_1)]\}}$. A uniform prior on c is a typical choice because this y-shift is related to the mean of observed data or the underlying process. The uniform prior on Δt constrains its values to ensure that the shifted light curves overlap in time. This naively informative hyperprior distribution on the parameters governing the underlying process is $p(\bar{M}, \sigma, \tau) \propto \tau^{-2} e^{-1/\tau}$, where $\bar{M}$, σ, τ are CAR parameters as defined above and in Paper I. This puts a uniform prior on $\bar{M}$ and σ, and an inverse-$\Gamma(1, 1)$ prior on τ.

The full posterior distribution was obtained by multiplying together (1) the likelihood for the state-space representation, (2) the prior for the underlying process, Δt, and c, and (3) the hyperpriors for $\bar{M}$, σ, and τ. The team proposed a Gibbs sampler for this full posterior distribution (algorithm 2) and its approximation (algorithm 1) in TDC1. Details of the two samplers were submitted to the "Evil" Team and will appear in a separate paper, in preparation. In order to obtain the time delay from its posterior distribution, three Markov chains were combined with starting values chosen randomly around the most likely values. Rigorous convergence checks of the Markov chains were conducted using trace plots, autocorrelation plots, and the Gelman–Rubin diagnostic statistic, applied to all of the model parameters.

The model did not account for the microlensing. However, when it was suspected it after a visual inspection, this team accounted for its polynomial long-term effect (linear or quadratic) by the regression and ran the model on the residuals. This worked well because the intrinsic variability of quasar data did not disappear even after the long-term trend was removed.

4. ANALYSIS OF THE SUBMISSIONS

4.1. *Lessons from TDC0 Applied to TDC1*

During the analysis of the TDC0 submissions, the "Evil" Team noticed that several teams were affected by outliers: most of their submitted time delay estimates were good, but a few differed from the truth by more than would be expected, given their uncertainties. To characterize this, an additional metric X was introduced: X is the fraction of pairs with $\chi_i^2 < 10$, i.e., the fraction without outliers. $X = 1$ means that none of the submitted delays is an outlier. Outliers in this category could stem from underestimated error bars, or for example by convergence on the wrong solution in the presence of light curve features (due to, e.g., microlensing) that are not taken into account by the method's model.

We will return to the issue of outliers, and how they can be identified based on lensing geometry or cosmological analysis, after we present the main results of TDC1. In this section, we give the unfiltered statistics as well as the metrics calculated after points with $\chi_i^2 > 10$ have been removed, in order to give an idea of how well a method *could* do if outliers could be identified and rejected.

We also consider an additional cut, based only on the accuracy parameter $|A_i| < 0.1$, and the related quantity X_A, which counts the fraction of systems satisfying this alternative criterion, i.e., we take $|A_i| > 0.1$ as outliers rather than $\chi_i^2 > 10$ in this case. This cut was chosen to quantify the number of systems for which the time-delay would be much more uncertain than the 3%–5% modeling error that can be obtained in the reconstruction of the difference in gravitational potential between two images in the best cases (Suyu et al. 2013, 2014). In some sense this cut filters out the systems that are not cosmologically consistent and thus could be rejected by a joint cosmological analysis.

Finally, as a third way to illustrate the potential of each method once outliers have been removed, we also consider the median, 16 and 84 percentile of the statistics A_i, P_i, and χ_i^2 for each method, as opposed to the means defined in Section 1.

4.2. Blind and Non-blind Submissions

One of the main goals of this TDC is to achieve a true blind testing of the algorithms. To achieve this, TDC0 truth files were not revealed until after the deadline of TDC1, lest they give too much away about the data generation process. In addition, upon requests from each "Good" Team we provided only minimal feedback after each submission, in the form of the metrics listed above rounded to two significant digits. This was deemed to be a reasonable compromise between preserving the blindness of the challenge, and helping teams to identify coding errors that had nothing to do with their actual chosen algorithms. Only submissions made prior to any feedback were considered truly blind, even though resubmissions by the teams who decided to take advantage of this opportunity were accepted. Resubmissions were considered not fully blind for the purpose of this analysis. Note that all of the "representative" submissions referred in later sections were made fully blind.

4.3. Basic Statistics

The metrics for each submission are shown in Tables 4 and 5, separated by challenge rung. In order to visually compare the different algorithms in a relatively clear manner, we have chosen to show only one submission for each team. This "representative" algorithm was chosen by each team after the true time delays were unblinded, and therefore it is somewhat indicative of the best performance of each method. Results for all the other submissions are available at the TDC Web site. Importantly, it should be kept in mind that this is a multi-dimensional problem, and there is not necessarily a "best" submission, not even within each method. Rather, each submission is a tradeoff between competing needs of achieving low P and A, while keeping χ^2 reasonable and f and X as high as possible. Some of the statistics are mathematically interdependent. For example, χ^2 and P both contain the submitted uncertainty estimates: teams could decide to reduce their χ^2 at the price of increasing their P, and vice versa.

The metrics obtained by these submissions are plotted in Figures 5–9. The plots show the metrics that have been computed directly from the submitted values, together with the recomputed metrics after rejecting the outliers using the $\chi_i^2 < 10$ cut. The corner plots in Figures 5–9 also show a shaded region that represents the TDC1 soft targets that were estimated in Paper I as the metric values needed for methods to be competitive, namely:

1. $f > 0.5$
2. $\chi^2 < 1.5$
3. $|A| < 0.03$ [goal 0.002]
4. $P < 0.03$.

As discussed in Paper I, in this exploratory challenge, these targets were deemed sufficient given the current lensed quasar sample of a few tens of systems. In the long run, for samples of thousands of lenses, a desirable goal is to improve the accuracy or bias to sub-percent level ($|A| < 0.2\%$; see Hojjati & Linder 2014 for the cosmological requirement derivation), such that the contribution of time delay measurement to the error budget of cosmological parameters would be smaller than the projected statistical uncertainties. We emphasize that these targets are approximate and only with a fully cosmological challenge would they be translated into a single indicator of performance, as we outline in the final section of this paper.

As is shown in the figures, most of the algorithms achieved the $|A|$ and χ^2 criteria, especially after the rejection of outliers in the submissions. The "Evil" Team's baseline method had a large fraction of outliers, but once those were rejected, it did not perform significantly worse than many of the "Good" Teams submissions. The criterion that proved more difficult to meet was the one on the success fraction f, where teams were typically closer to the threshold for TDC0 (shown also in the cornerplot as a lighter shaded region) than for TDC1. As we discuss below, this is due to the strategy that most teams followed, i.e., to have high standards of acceptance in order to reduce outliers. Notably, for many of the methods $|A|$ is at the sub-percent level—well below the target of 0.03—which is very promising in view of future cosmological studies.

Interestingly, the "Evil" light curves did not yield significantly poorer statistics than the regular ones. From this comparison we infer that the methods used are generally robust to small and realistic unknown light curve systematics like the ones introduced by the "Evil" Team. This is encouraging and bodes well for the application of the methods to real data.

4.4. Trends with Intrinsic Properties of the Light Curves and Implications for Future Work

We now investigate how the quality of the inferred time delays depends on the intrinsic properties of the light curves. We wish to discover general trends that are not inherent to the peculiarities of each method. In order to carry out this investigation, in Figure 10 we plot the individual accuracy, precision and goodness of fit of each submission (A_i, P_i, and χ_i^2) as a function of true time delay, the variability parameters of the intrinsic quasar light curves (τ and σ), and the magnitude of the fainter image of each pair (i_2). In this illustration we show the results for Rung 1; the other rungs give similar results. Figure 11 shows summary statistics of the same data, represented by the average statistics in bins of the variable on the abscissa—the color scheme is the same as described in the legend to Figure 5.

We can see in these figures a few global trends. The most prominent appears to be between P and the true time delay. P decreases with time delay consistent with the time delay uncertainty being approximately constant in days, as expected if the absolute precision is driven by the sampling of the light curves. Qualitatively, P_i and A_i also appear to decrease (i.e., improve) as σ increases, also as expected: the light curves with the highest variability amplitudes should be easier to interpret and therefore should yield higher precision and fewer outliers.

Remarkably, we see very little dependency on i_2, as if the signal-to-noise ratio of the fainter image is not as important, once it is passes some minimum threshold. This suggests that the simulated data are of sufficient quality and that the photometric uncertainty is subdominant with respect to the uncertainties introduced by microlensing and sampling. The weak dependency on the magnitude of the fainter image i_2 implies that the statistics we derive from the TDC1 data set are very similar to what we would have derived from a random subset of OM10. In fact, by recomputing weighted averages of the statistics to match the OM10 i_2 magnitude distribution, we verified that the changes of the statistics would have been comparable to their uncertainty.

Finally, we investigated the level of agreement between the algorithms to see whether success was due solely to the properties of the light curves or whether it depended on the specifics of each algorithm. The results are shown in Figure 12 for three

Table 4

Mean and Median Statistics for the "Representative" Submissions

Method	Rung	f	χ^2	P	A	χ^2_{median}	P_{median}	A_{median}
0	0	0.36	195000 ± 76000	0.078 ± 0.004	-0.181 ± 0.065	$0.085^{189}_{0.078}$	$0.055^{0.083}_{0.036}$	$-0.004^{0.025}_{0.86}$
0	1	0.36	390000 ± 150000	0.08 ± 0.005	-0.281 ± 0.061	$0.47^{2046}_{0.46}$	$0.052^{0.088}_{0.039}$	$-0.021^{0.04}_{0.98}$
0	2	0.32	3996 ± 1052	0.082 ± 0.005	-0.28 ± 0.042	$0.42^{1199}_{0.4}$	$0.059^{0.088}_{0.041}$	$-0.02^{0.05}_{0.97}$
0	3	0.33	920000 ± 500000	0.08 ± 0.005	-0.247 ± 0.053	$0.37^{2527}_{0.36}$	$0.05^{0.098}_{0.036}$	$-0.013^{0.034}_{0.97}$
0	4	0.35	950000 ± 240000	0.042 ± 0.004	-0.712 ± 0.03	16136^{671657}_{16136}	$0.008^{0.087}_{0.007}$	$-1.0^{0.99}_{0.007}$
1	0	0.53	0.579 ± 0.047	0.038 ± 0.001	-0.018 ± 0.001	$0.26^{0.77}_{0.22}$	$0.034^{0.028}_{0.016}$	$-0.015^{0.016}_{0.024}$
1	1	0.37	0.543 ± 0.049	0.045 ± 0.001	-0.022 ± 0.001	$0.24^{0.69}_{0.22}$	$0.04^{0.025}_{0.015}$	$-0.02^{0.017}_{0.022}$
1	2	0.35	0.89 ± 0.19	0.053 ± 0.001	-0.025 ± 0.002	$0.23^{0.92}_{0.21}$	$0.047^{0.034}_{0.021}$	$-0.02^{0.024}_{0.038}$
1	3	0.34	0.524 ± 0.077	0.059 ± 0.002	-0.021 ± 0.002	$0.17^{0.67}_{0.15}$	$0.051^{0.037}_{0.02}$	$-0.018^{0.025}_{0.029}$
1	4	0.35	0.608 ± 0.072	0.056 ± 0.002	-0.024 ± 0.002	$0.2^{0.84}_{0.18}$	$0.051^{0.036}_{0.024}$	$-0.019^{0.024}_{0.035}$
2	0	0.53	0.125 ± 0.011	0.205 ± 0.007	-0.017 ± 0.004	$0.043^{0.178}_{0.039}$	$0.151^{0.198}_{0.078}$	$-0.008^{0.046}_{0.062}$
2	1	0.27	0.138 ± 0.016	0.233 ± 0.01	-0.025 ± 0.006	$0.054^{0.216}_{0.05}$	$0.19^{0.17}_{0.1}$	$-0.008^{0.05}_{0.086}$
2	2	0.21	0.043 ± 0.004	0.242 ± 0.01	-0.015 ± 0.004	$0.021^{0.058}_{0.019}$	$0.201^{0.207}_{0.092}$	$-0.009^{0.04}_{0.056}$
2	3	0.3	0.099 ± 0.013	0.247 ± 0.011	-0.03 ± 0.006	$0.039^{0.121}_{0.035}$	$0.17^{0.266}_{0.085}$	$-0.013^{0.046}_{0.08}$
2	4	0.21	0.178 ± 0.018	0.363 ± 0.015	-0.059 ± 0.011	$0.097^{0.252}_{0.084}$	$0.32^{0.27}_{0.15}$	$-0.04^{0.12}_{0.15}$
3	0	0.53	1.068 ± 0.069	0.043 ± 0.003	-0.0 ± 0.003	$0.46^{1.67}_{0.4}$	$0.022^{0.041}_{0.012}$	$0.0^{0.025}_{0.025}$
3	1	0.26	1.031 ± 0.097	0.04 ± 0.003	0.008 ± 0.003	$0.49^{1.47}_{0.46}$	$0.027^{0.034}_{0.014}$	$0.004^{0.033}_{0.026}$
3	2	0.21	1.02 ± 0.13	0.043 ± 0.004	-0.002 ± 0.004	$0.38^{1.43}_{0.34}$	$0.026^{0.037}_{0.013}$	$0.003^{0.02}_{0.033}$
3	3	0.3	0.813 ± 0.074	0.068 ± 0.006	-0.004 ± 0.006	$0.39^{1.04}_{0.37}$	$0.034^{0.066}_{0.019}$	$-0.002^{0.032}_{0.032}$
3	4	0.21	1.07 ± 0.23	0.098 ± 0.014	0.0 ± 0.008	$0.24^{1.41}_{0.22}$	$0.064^{0.06}_{0.034}$	$0.003^{0.054}_{0.04}$
4	0	0.53	0.497 ± 0.047	0.033 ± 0.002	-0.0 ± 0.001	$0.15^{0.75}_{0.14}$	$0.018^{0.038}_{0.011}$	$0.0^{0.012}_{0.012}$
4	1	0.27	0.528 ± 0.066	0.028 ± 0.002	0.0 ± 0.002	$0.16^{0.78}_{0.15}$	$0.02^{0.021}_{0.01}$	$-0.001^{0.015}_{0.012}$
4	2	0.21	0.464 ± 0.069	0.028 ± 0.002	-0.001 ± 0.002	$0.15^{0.54}_{0.13}$	$0.02^{0.023}_{0.011}$	$0.0^{0.013}_{0.01}$
4	3	0.3	0.542 ± 0.074	0.042 ± 0.003	-0.003 ± 0.003	$0.16^{0.76}_{0.14}$	$0.023^{0.038}_{0.013}$	$-0.001^{0.017}_{0.015}$
4	4	0.21	0.665 ± 0.065	0.045 ± 0.003	0.001 ± 0.003	$0.31^{0.94}_{0.29}$	$0.032^{0.035}_{0.015}$	$-0.001^{0.035}_{0.028}$
5	0	0.68	0.91 ± 0.092	0.032 ± 0.001	0.003 ± 0.002	$0.24^{1.19}_{0.23}$	$0.024^{0.034}_{0.014}$	$0.001^{0.022}_{0.015}$
5	1	0.27	1.76 ± 0.42	0.037 ± 0.002	-0.002 ± 0.003	$0.39^{1.86}_{0.36}$	$0.03^{0.029}_{0.015}$	$-0.001^{0.026}_{0.026}$
5	2	0.32	1.57 ± 0.21	0.043 ± 0.001	-0.003 ± 0.004	$0.44^{1.93}_{0.41}$	$0.036^{0.036}_{0.017}$	$-0.001^{0.036}_{0.043}$
5	3	0.35	1.89 ± 0.31	0.036 ± 0.001	0.002 ± 0.003	$0.42^{2.3}_{0.4}$	$0.029^{0.03}_{0.015}$	$0.001^{0.029}_{0.031}$
5	4	0.18	7.2 ± 2.7	0.05 ± 0.003	-0.021 ± 0.007	$1.5^{4.3}_{1.4}$	$0.043^{0.04}_{0.021}$	$-0.016^{0.072}_{0.068}$
6	0	0.04	0.32 ± 0.071	0.077 ± 0.017	0.005 ± 0.011	$0.11^{0.66}_{0.1}$	$0.044^{0.06}_{0.027}$	$0.0^{0.027}_{0.037}$
6	1	0.02	66 ± 64	0.175 ± 0.055	2.3 ± 2.2	$0.36^{0.25}_{0.27}$	$0.093^{0.13}_{0.037}$	$0.042^{0.056}_{0.047}$
6	2	0.03	0.71 ± 0.21	0.142 ± 0.021	0.027 ± 0.032	$0.37^{0.78}_{0.36}$	$0.117^{0.098}_{0.064}$	$0.029^{0.077}_{0.089}$
6	3	0.02	1.7 ± 1.2	0.168 ± 0.031	0.14 ± 0.1	$0.33^{0.9}_{0.3}$	$0.118^{0.079}_{0.068}$	$0.02^{0.119}_{0.056}$
6	4	0.01	0.19 ± 0.1	0.55 ± 0.12	0.169 ± 0.058	$0.066^{0.169}_{0.051}$	$0.48^{0.2}_{0.25}$	$0.16^{0.24}_{0.19}$
7	0	0.33	65 ± 51	0.04 ± 0.003	-0.011 ± 0.009	$0.6^{3.44}_{0.55}$	$0.021^{0.057}_{0.015}$	$-0.0^{0.029}_{0.034}$
7	1	0.24	2.71 ± 0.5	0.036 ± 0.003	0.002 ± 0.006	$0.67^{3.14}_{0.62}$	$0.021^{0.045}_{0.015}$	$0.001^{0.034}_{0.029}$
7	2	0.37	3.21 ± 0.55	0.04 ± 0.003	0.008 ± 0.006	$0.74^{3.39}_{0.69}$	$0.023^{0.051}_{0.015}$	$-0.0^{0.036}_{0.029}$
7	3	0.3	2.39 ± 0.39	0.051 ± 0.004	0.02 ± 0.012	$0.65^{2.49}_{0.6}$	$0.025^{0.067}_{0.018}$	$-0.0^{0.035}_{0.035}$
7	4	0.22	185 ± 119	0.062 ± 0.005	-0.03 ± 0.02	$0.6^{3.47}_{0.53}$	$0.035^{0.104}_{0.026}$	$-0.001^{0.061}_{0.064}$
8	0	0.44	109 ± 58	0.047 ± 0.004	-0.025 ± 0.032	$0.16^{1.21}_{0.15}$	$0.025^{0.047}_{0.016}$	$0.0^{0.019}_{0.021}$
8	1	0.22	88 ± 38	0.101 ± 0.05	-0.02 ± 0.019	$0.17^{2.4}_{0.16}$	$0.029^{0.066}_{0.016}$	$0.0^{0.026}_{0.04}$
8	2	0.18	91 ± 72	0.07 ± 0.006	-0.006 ± 0.019	$0.14^{0.81}_{0.14}$	$0.046^{0.076}_{0.028}$	$0.0^{0.032}_{0.032}$
8	3	0.19	27 ± 21	0.059 ± 0.004	-0.008 ± 0.013	$0.24^{1.34}_{0.23}$	$0.041^{0.064}_{0.025}$	$0.001^{0.033}_{0.041}$
8	4	0.16	2.6 ± 1.1	0.068 ± 0.004	-0.0 ± 0.006	$0.3^{1.29}_{0.28}$	$0.055^{0.07}_{0.032}$	$0.001^{0.045}_{0.049}$
9	4	0.27	8.7 ± 3.5	0.035 ± 0.002	0.003 ± 0.006	$0.55^{2.79}_{0.47}$	$0.024^{0.037}_{0.014}$	$0.0^{0.031}_{0.042}$

Notes. Method: 0, Rumbaugh-Gold; 1, Shafieloo-Arman7; 2, PyCS-d3cs-vanilla-dou-full; 3, PyCS-sdi-vanilla-dou-full; 4, PyCS-spl-vanilla-dou-full; 5, Jackson-manchester2_0_3_4; 6, Kumar; 7, JPL; 8, Hojjati-Stark; 9, DeltaTBayes-DeltaTBayes1.

representative rungs. Clearly, some light curves do not contain enough information for any method to be successful (hence the peak at zero). In Rung 0, there is a bump around 6 indicating that for very good light curve a majority of the methods are successful. However, as the quality of data degrades in the next rungs it appears that there is a continuum distribution. Therefore we conclude that different methods pick up different features of the light curves and accuracy varies widely between methods.

Table 5
Filtered Statistics for the "Representative" Submissions

Method	Rung	$f_{3.3\sigma}$	$\chi^2_{3.3\sigma}$	$P_{3.3\sigma}$	$A_{3.3\sigma}$	X	f_A	χ^2_A	P_A	A_A	X_A
0	0	0.29	0.379 ± 0.072	0.087 ± 0.005	−0.003 ± 0.004	0.8	0.28	0.299 ± 0.056	0.08 ± 0.004	−0.0 ± 0.002	0.77
0	1	0.23	0.577 ± 0.095	0.096 ± 0.006	−0.01 ± 0.007	0.65	0.22	3.9 ± 2.3	0.082 ± 0.005	−0.004 ± 0.002	0.62
0	2	0.23	0.8 ± 0.11	0.098 ± 0.005	−0.007 ± 0.006	0.73	0.21	0.74 ± 0.23	0.09 ± 0.005	−0.002 ± 0.003	0.66
0	3	0.22	0.59 ± 0.1	0.097 ± 0.006	0.0 ± 0.007	0.66	0.21	1.26 ± 0.4	0.087 ± 0.006	−0.002 ± 0.002	0.64
0	4	0.11	0.37 ± 0.11	0.119 ± 0.009	−0.009 ± 0.006	0.3	0.1	0.26 ± 0.058	0.112 ± 0.008	−0.003 ± 0.004	0.28
1	0	0.53	0.552 ± 0.04	0.038 ± 0.001	−0.017 ± 0.001	1.0	0.52	0.53 ± 0.038	0.038 ± 0.001	−0.017 ± 0.001	0.99
1	1	0.37	0.543 ± 0.049	0.045 ± 0.001	−0.022 ± 0.001	1.0	0.36	0.497 ± 0.041	0.044 ± 0.001	−0.021 ± 0.001	0.99
1	2	0.35	0.673 ± 0.068	0.053 ± 0.001	−0.025 ± 0.002	0.99	0.33	0.73 ± 0.19	0.052 ± 0.001	−0.021 ± 0.002	0.95
1	3	0.34	0.458 ± 0.039	0.059 ± 0.002	−0.02 ± 0.002	1.0	0.33	0.419 ± 0.036	0.058 ± 0.002	−0.018 ± 0.002	0.97
1	4	0.35	0.559 ± 0.052	0.056 ± 0.002	−0.024 ± 0.002	1.0	0.33	0.535 ± 0.069	0.055 ± 0.002	−0.021 ± 0.002	0.97
2	0	0.53	0.125 ± 0.011	0.205 ± 0.007	−0.017 ± 0.004	1.0	0.45	0.081 ± 0.008	0.17 ± 0.006	−0.005 ± 0.002	0.83
2	1	0.27	0.138 ± 0.016	0.233 ± 0.01	−0.025 ± 0.006	1.0	0.21	0.078 ± 0.01	0.191 ± 0.008	−0.006 ± 0.003	0.79
2	2	0.21	0.043 ± 0.004	0.242 ± 0.01	−0.015 ± 0.004	1.0	0.19	0.033 ± 0.004	0.217 ± 0.009	−0.007 ± 0.003	0.9
2	3	0.3	0.099 ± 0.013	0.247 ± 0.011	−0.03 ± 0.006	1.0	0.25	0.056 ± 0.005	0.201 ± 0.01	−0.007 ± 0.003	0.83
2	4	0.21	0.178 ± 0.018	0.363 ± 0.015	−0.059 ± 0.011	1.0	0.12	0.063 ± 0.008	0.287 ± 0.018	−0.007 ± 0.005	0.55
3	0	0.53	1.048 ± 0.066	0.043 ± 0.003	−0.0 ± 0.003	1.0	0.5	0.956 ± 0.068	0.037 ± 0.003	0.001 ± 0.001	0.94
3	1	0.26	0.977 ± 0.081	0.04 ± 0.003	0.006 ± 0.003	1.0	0.25	0.858 ± 0.069	0.037 ± 0.003	0.004 ± 0.002	0.95
3	2	0.21	0.94 ± 0.1	0.043 ± 0.004	−0.002 ± 0.004	0.99	0.2	0.92 ± 0.13	0.035 ± 0.002	−0.003 ± 0.002	0.93
3	3	0.3	0.813 ± 0.074	0.068 ± 0.006	−0.004 ± 0.006	1.0	0.27	0.747 ± 0.073	0.05 ± 0.004	−0.003 ± 0.002	0.92
3	4	0.21	0.804 ± 0.096	0.098 ± 0.015	0.005 ± 0.006	0.99	0.19	0.64 ± 0.11	0.069 ± 0.004	0.005 ± 0.003	0.86
4	0	0.53	0.472 ± 0.04	0.033 ± 0.002	−0.0 ± 0.001	1.0	0.52	0.483 ± 0.048	0.029 ± 0.001	0.0 ± 0.001	0.98
4	1	0.27	0.528 ± 0.066	0.028 ± 0.002	0.0 ± 0.002	1.0	0.27	0.467 ± 0.051	0.027 ± 0.002	−0.0 ± 0.001	0.99
4	2	0.21	0.464 ± 0.069	0.028 ± 0.002	−0.001 ± 0.002	1.0	0.21	0.431 ± 0.064	0.028 ± 0.002	−0.001 ± 0.001	0.99
4	3	0.3	0.494 ± 0.057	0.042 ± 0.003	−0.001 ± 0.003	1.0	0.29	0.455 ± 0.052	0.037 ± 0.003	−0.001 ± 0.001	0.97
4	4	0.21	0.665 ± 0.065	0.045 ± 0.003	0.001 ± 0.003	1.0	0.2	0.571 ± 0.056	0.041 ± 0.002	0.0 ± 0.002	0.95
5	0	0.68	0.741 ± 0.053	0.032 ± 0.001	0.004 ± 0.002	0.99	0.65	0.659 ± 0.054	0.03 ± 0.001	0.002 ± 0.001	0.95
5	1	0.27	0.926 ± 0.098	0.037 ± 0.002	−0.003 ± 0.003	0.97	0.26	1.42 ± 0.42	0.034 ± 0.002	−0.001 ± 0.002	0.93
5	2	0.31	1.083 ± 0.096	0.043 ± 0.001	−0.002 ± 0.003	0.97	0.29	1.08 ± 0.13	0.04 ± 0.001	−0.001 ± 0.002	0.92
5	3	0.34	1.165 ± 0.099	0.036 ± 0.001	0.002 ± 0.003	0.98	0.32	1.23 ± 0.17	0.032 ± 0.001	0.0 ± 0.002	0.91
5	4	0.16	2.12 ± 0.2	0.052 ± 0.003	−0.015 ± 0.007	0.92	0.15	5.4 ± 3.1	0.044 ± 0.002	−0.011 ± 0.004	0.82
6	0	0.04	0.32 ± 0.071	0.077 ± 0.017	0.005 ± 0.011	1.0	0.04	0.32 ± 0.073	0.063 ± 0.01	−0.004 ± 0.006	0.97
6	1	0.02	0.334 ± 0.051	0.121 ± 0.016	0.04 ± 0.014	0.95	0.02	0.31 ± 0.053	0.111 ± 0.016	0.027 ± 0.012	0.86
6	2	0.03	0.71 ± 0.21	0.142 ± 0.021	0.027 ± 0.032	1.0	0.02	0.333 ± 0.087	0.111 ± 0.012	0.019 ± 0.011	0.75
6	3	0.02	0.51 ± 0.15	0.155 ± 0.03	0.037 ± 0.02	0.95	0.02	0.278 ± 0.095	0.13 ± 0.034	−0.003 ± 0.011	0.64
6	4	0.01	0.19 ± 0.1	0.55 ± 0.12	0.169 ± 0.058	1.0	0.0	0.024 ± 0.011	0.358 ± 0.075	−0.005 ± 0.026	0.33
7	0	0.31	1.42 ± 0.12	0.041 ± 0.003	−0.001 ± 0.004	0.95	0.3	1.82 ± 0.28	0.033 ± 0.003	−0.001 ± 0.002	0.89
7	1	0.23	1.39 ± 0.13	0.037 ± 0.003	−0.0 ± 0.006	0.95	0.22	2.25 ± 0.47	0.028 ± 0.002	0.002 ± 0.002	0.91
7	2	0.35	1.41 ± 0.1	0.04 ± 0.003	0.006 ± 0.004	0.94	0.33	2.06 ± 0.34	0.032 ± 0.002	−0.001 ± 0.002	0.89
7	3	0.28	1.28 ± 0.11	0.051 ± 0.004	0.007 ± 0.007	0.95	0.26	1.82 ± 0.32	0.033 ± 0.002	−0.003 ± 0.002	0.87
7	4	0.21	1.33 ± 0.14	0.063 ± 0.005	0.003 ± 0.007	0.93	0.18	1.93 ± 0.44	0.043 ± 0.004	0.002 ± 0.003	0.79
8	0	0.42	0.531 ± 0.054	0.047 ± 0.004	−0.0 ± 0.002	0.95	0.41	0.81 ± 0.14	0.041 ± 0.003	−0.001 ± 0.001	0.93
8	1	0.2	0.596 ± 0.087	0.105 ± 0.056	−0.004 ± 0.004	0.9	0.2	0.76 ± 0.14	0.101 ± 0.057	−0.001 ± 0.002	0.89
8	2	0.17	0.62 ± 0.11	0.07 ± 0.006	0.003 ± 0.004	0.96	0.16	0.354 ± 0.064	0.064 ± 0.005	−0.001 ± 0.003	0.88
8	3	0.18	0.78 ± 0.12	0.06 ± 0.004	−0.003 ± 0.005	0.96	0.17	1.03 ± 0.34	0.053 ± 0.004	0.0 ± 0.003	0.89
8	4	0.16	0.89 ± 0.14	0.07 ± 0.004	0.002 ± 0.005	0.98	0.15	1.59 ± 0.69	0.063 ± 0.004	0.002 ± 0.003	0.9
9	4	0.25	1.2 ± 0.1	0.036 ± 0.003	−0.006 ± 0.004	0.94	0.25	3.7 ± 1.4	0.03 ± 0.002	−0.002 ± 0.002	0.91

Notes. Method 0:Rumbaugh-Gold, 1:Shafieloo-Arman7, 2:PyCS-d3cs-vanilla-dou-full, 3:PyCS-sdi-vanilla-dou-full, 4:PyCS-spl-vanilla-dou-full, 5:Jackson-manchester2_0_3_4, 6:Kumar, 7:JPL, 8:Hojjati-Stark, 9:DeltaTBayes-DeltaTBayes1.

5. IMPLICATIONS FOR OBSERVING STRATEGY

By comparing the results from the different TDC1 rungs, we can now answer the following question: How does time delay measurement accuracy depend on observing cadence, season length and campaign length?

Figure 13 shows the variation in the $|A|$, P, and f metrics with cadence and season length, assuming outliers to have been rejected by $|A_i| > 0.1$. Each pair of connected points plotted in the panels of this figure represents a simple test where the control variable (cadence or season length) is varied, while keeping the others constant. Campaign length and cadence regularity were also investigated in a similar manner, but the results—which are less striking—are not shown here. The six tests we carried out in total are summarized in Table 6. The top two rows in the table correspond to the plots shown in the left and right columns of the figure, respectively.

Figure 13 shows some interesting diversity between methods. Despite this, some approximate general trends can be seen. Greater accuracy and success fractions seem to be associated primarily with longer seasons, but there is considerable scatter between submissions, perhaps due to residual outliers in some cases. In most methods, little dependence of accuracy on cadence, campaign lengths beyond five years, or the regularity of the sampling was seen. The success fraction seems to be somewhat dependent on cadence but less so on campaign length. In general, the trends in precision with cadence and season length seem to be less marked, and show less scatter, than those

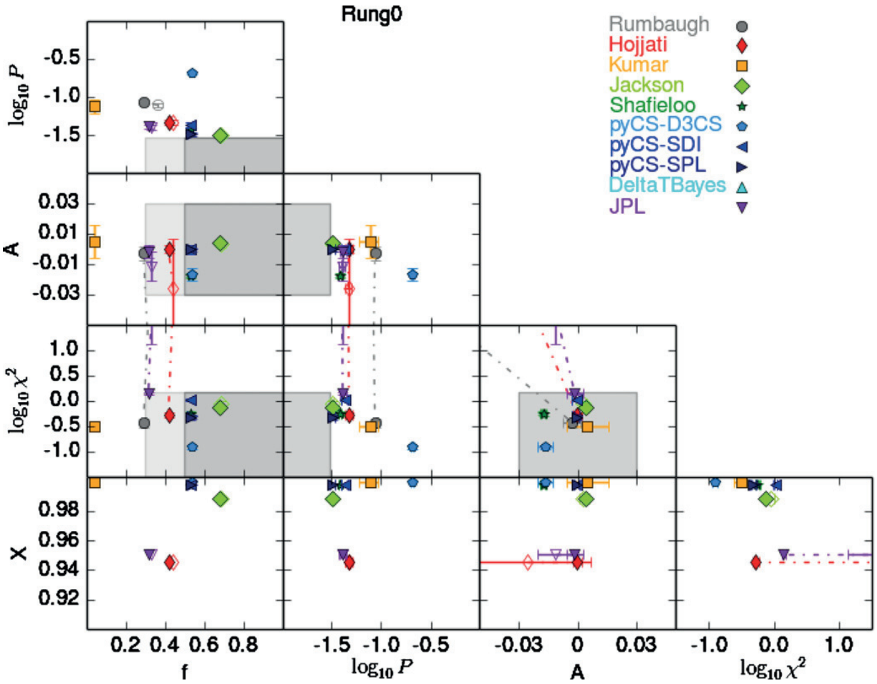


Figure 5. Results for TDC1 Rung 0, showing metrics for the "representative" submission for each of the 10 algorithms. This includes the baseline submission by the "Evil" Team ("Rumbaugh"). The f, P, A, and χ^2 metrics are defined in Section 1, while X is defined in Section 4.1. The shaded regions of each plot represent the soft targets for TDC1, as presented in the TDC0 paper. Both unfiltered results (open symbols) and results filtered by $\chi_i^2 < 10$ (solid symbols) are presented, and they are connected by dashed lines to show the improvements. Rung 0 simulates 3 day cadence and 8 month seasons over a 5 yr campaign with 400 observations in total (Table 1).

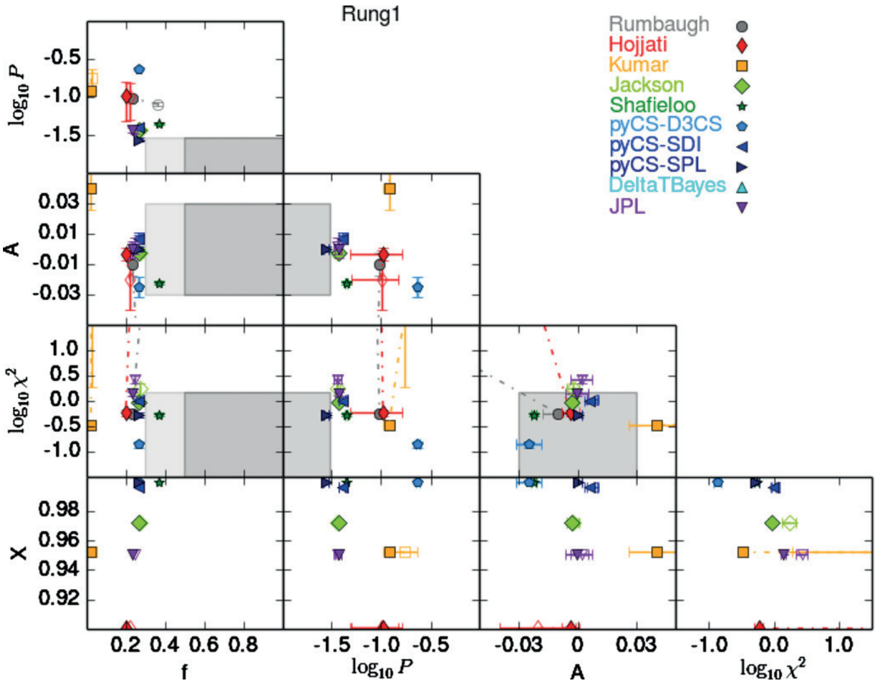


Figure 6. Same as Figure 5, but showing the results of TDC1 Rung 1, which simulates 3 day cadence and 4 month seasons over a 10 yr campaign with 400 observations in total (Table 1).

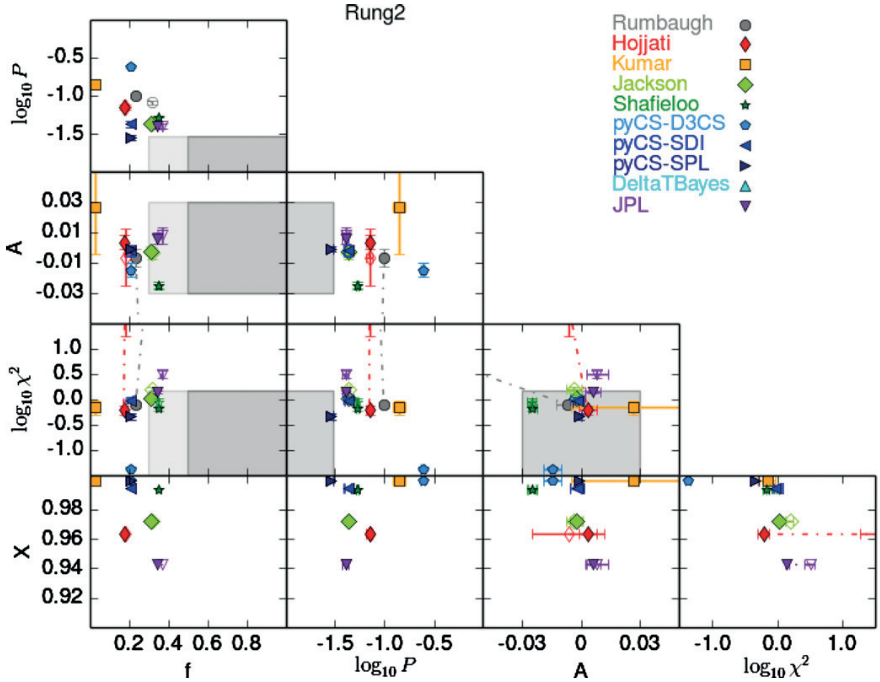


Figure 7. Same as Figure 5, but showing the results of TDC1 Rung 2, which simulates 3 day cadence and 4 month seasons over a 5 yr campaign with 200 observations in total (Table 1), and exactly regular time sampling.

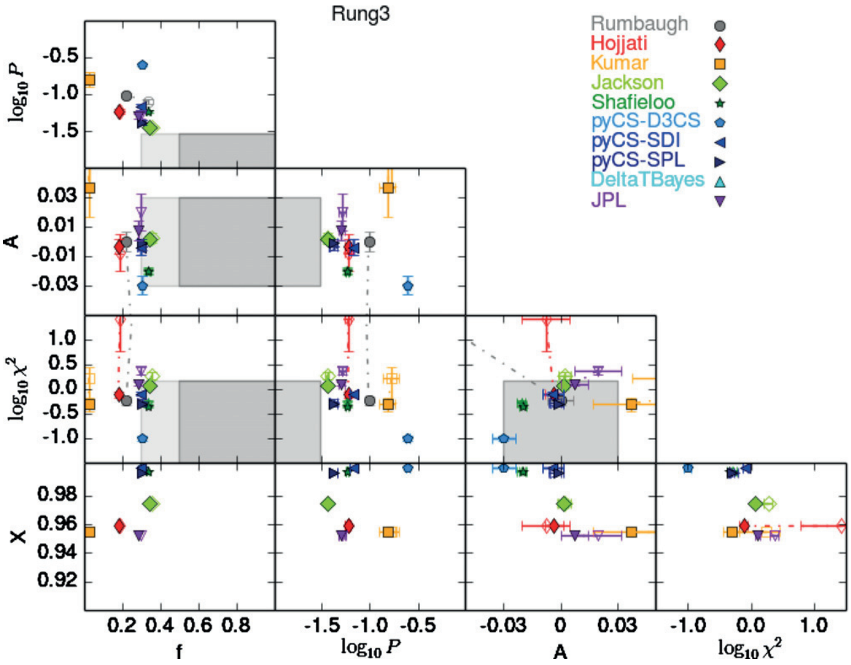


Figure 8. Same as Figure 5, but showing the results of TDC1, Rung 3, which simulates 3 day cadence and 4 month seasons over a 5 yr campaign with 200 observations in total (Table 1), and with 1 day scatter in the separations between observations.

in accuracy and success fraction. In general, cadence seems to be the most important factor for precision.

While the variation of time delay measurement with observing strategy seems to be somewhat algorithm-dependent, we can nevertheless hope to capture the general trends just described. Focusing on the PyCS-SPL results, we derived a very approximate power-law model for how the A, P, and f metrics varied with the main three quantities that describe the observing strategies in the rungs, mean cadence (cad), mean season length (sea), and campaign length (camp).

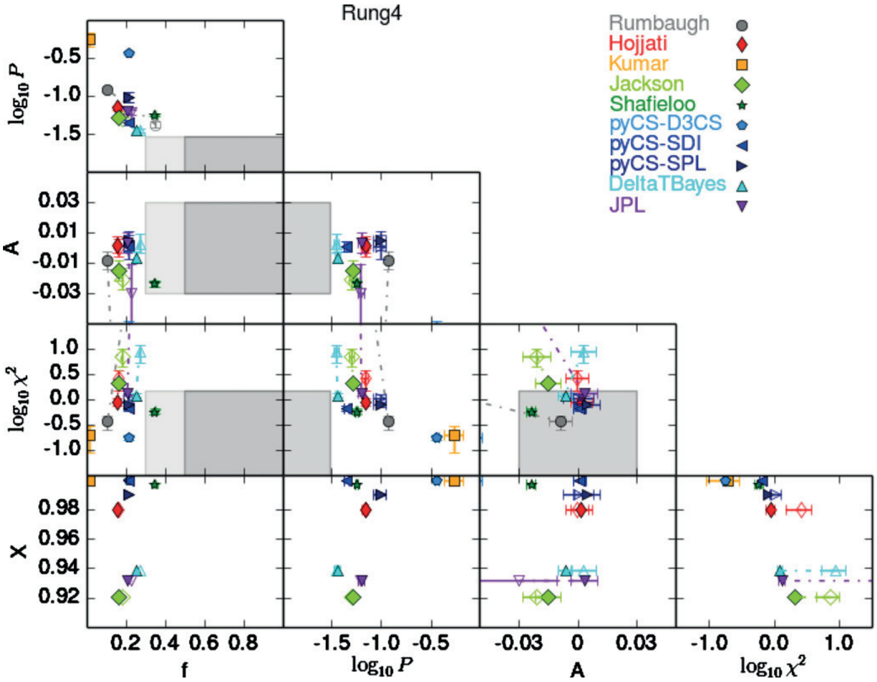


Figure 9. Same as Figure 5, but showing the results of TDC1, Rung 4, which simulates 6 day cadence and 4 month seasons over a 10 yr campaign with 200 observations in total (Table 1).

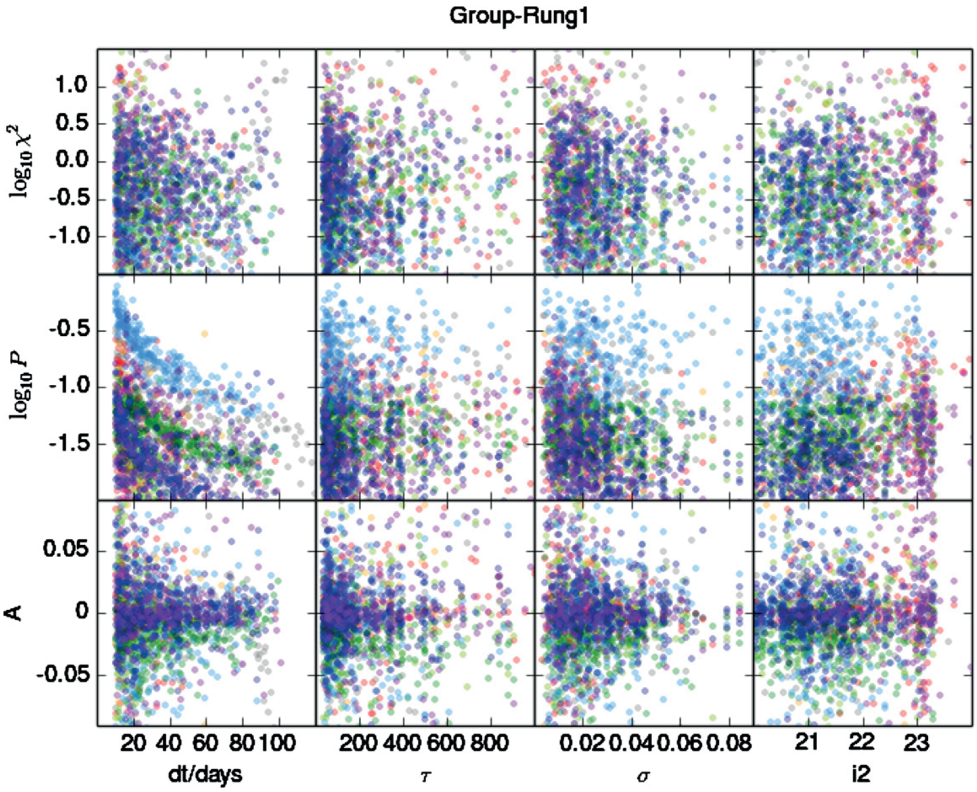


Figure 10. Unfiltered results of Rung 1: individual metrics of each "representative" submission (A_i, P_i, χ_i^2) as a function of true time delay dt, the variability parameters of the intrinsic quasar light curves (τ, σ), and the magnitude of the fainter image of each pair (i_2). The color scheme is the same as that described in the legend of Figure 5.

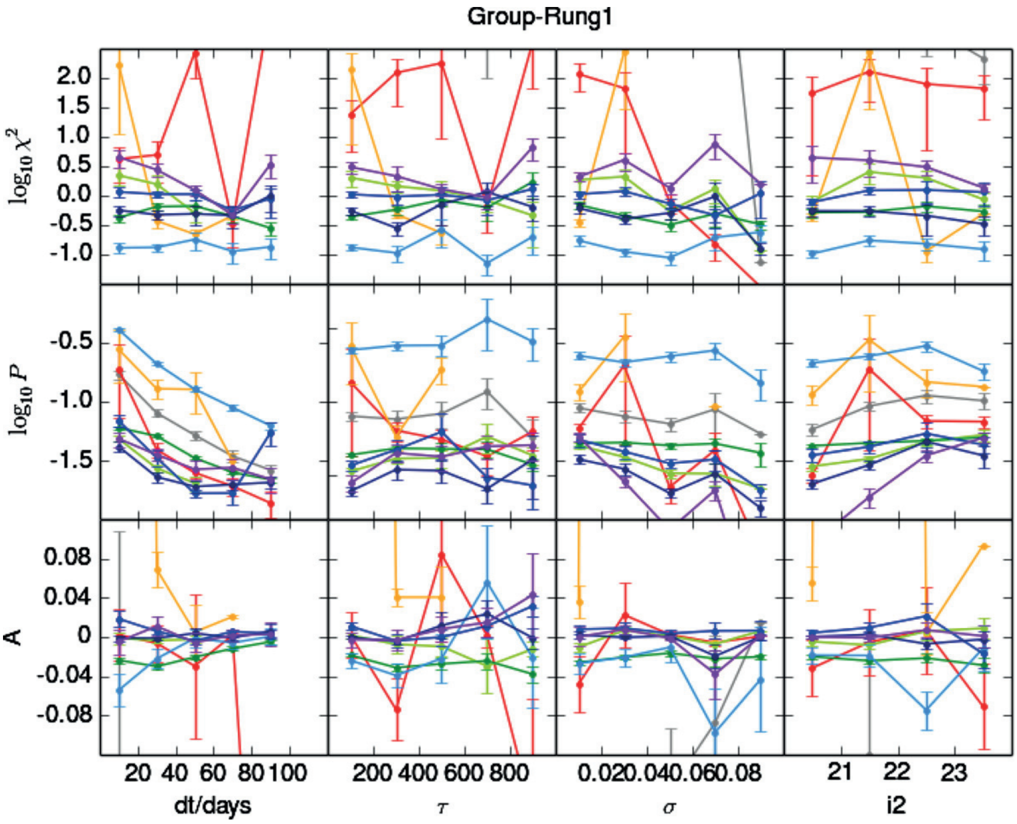


Figure 11. Summary unfiltered statistics of the same data in Figure 10, represented by the average statistics in bins of the variable on the abscissa. The color scheme is the same as described in the legend to Figure 5.

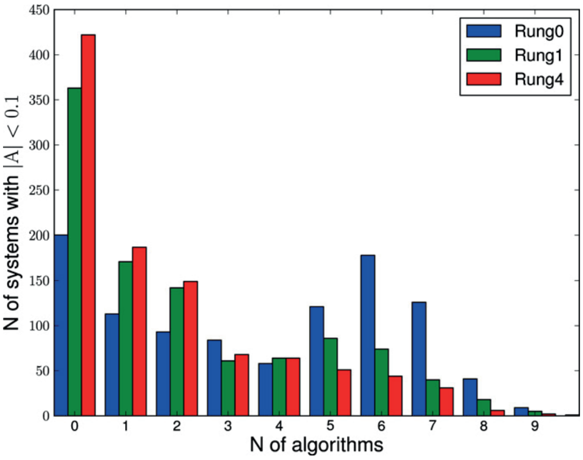


Figure 12. Distribution of the number of systems for which the time delay was successfully measured to a level of $|A_i| < 0.1$, plotted as a function of the number of algorithms (out of 10) that measure the time delays to this level. The plot shows Rungs 0, 1, and 4, which represent COSMOGRAIL-like, "optimistic" LSST, and "realistic" LSST programs, respectively. For Rung 0, there were more than ~200 systems for which none of the algorithms achieved the desired A, but also a large number of systems for which five, six, or seven of the algorithms successfully recovered this level. For Rungs 1 and 4, fewer of the systems were successfully recovered at the $|A_i| < 0.1$ level.

We find

$$|A|_{\rm model} \approx 0.06\% \left(\frac{\rm cad}{3\,\rm days}\right)^{0.0} \left(\frac{\rm sea}{4\,\rm months}\right)^{-1.0} \left(\frac{\rm camp}{5\,\rm yr}\right)^{-1.1}$$

$$P_{\rm model} \approx 4.0\% \left(\frac{\rm cad}{3\,\rm days}\right)^{0.7} \left(\frac{\rm sea}{4\,\rm months}\right)^{-0.3} \left(\frac{\rm camp}{5\,\rm yr}\right)^{-0.6}$$

$$f_{\rm model} \approx 30\% \left(\frac{\rm cad}{3\,\rm days}\right)^{-0.4} \left(\frac{\rm sea}{4\,\rm months}\right)^{0.8} \left(\frac{\rm camp}{5\,\rm yr}\right)^{-0.2}.$$

We can see that in this model, the accuracy metric A is the most sensitive to the observing strategy. It is also the case that it is the metric most sensitive to how the outliers are rejected. Rejecting outliers that have $\chi_i^2 > 10$ gives similar conclusions to those drawn here, but slightly different model parameters, in the sense that there is even stronger dependence of A on the observing strategy. In both cases the dependence of A on cadence is relatively weak. The season length and campaign length seem to be more important parameters: doubling either of these results in approximately a factor of two improvement in A. We note that constraining the total number of observations weakens these dependencies somewhat: for example, at fixed cadence, lengthening the season means shortening the campaign, and in our model, $|A|$ then decreases only as the ratio of the season length to the campaign length to the power of 0.1. The results of the fixed epoch number tests in Table 6 bore this out.

The precision and success fraction metrics' dependence on observing strategy is weaker, but it is interesting to note that

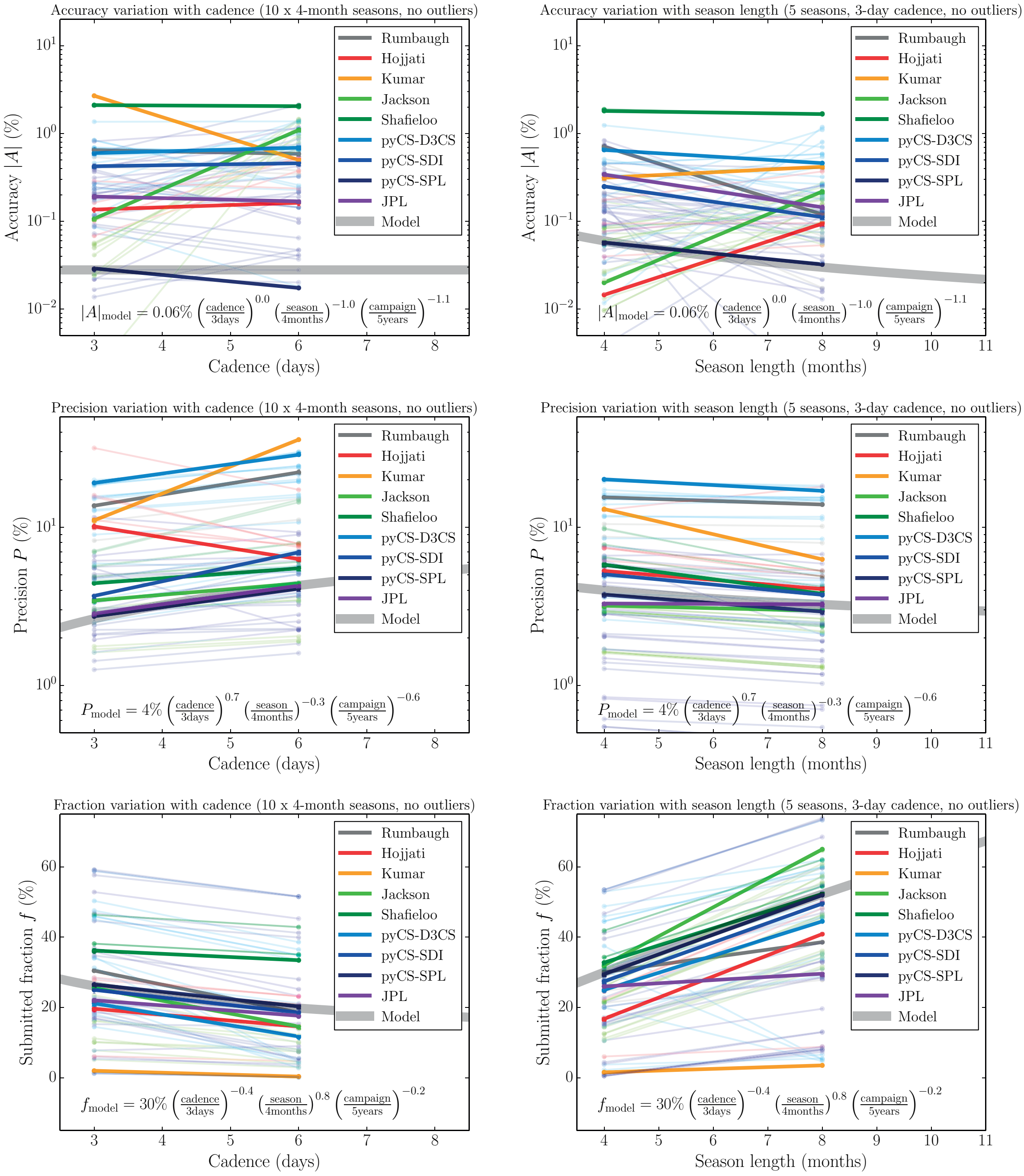


Figure 13. Changes in accuracy A (top row), precision P (middle row) and success fraction f (bottom row) with cadence (left) and season length (right), seen in the different TDC1 submissions. The gray approximate power-law model was derived by visual inspection of the pyCS-SPL results; the signs of the indices were pre-determined according to our expectations.

the precision depends more strongly on cadence than the season length, while the opposite is true for the success fraction. This can be understood qualitatively as the presence of large gaps reducing the overlap between light curves, making it more difficult to reliably and uniquely identify common features between them. Conversely, if the signal is properly identified, then the precision is driven by the total number of observation points, i.e., a combination of cadence and campaign duration. As a rough rule of thumb, we might have in mind that season length largely determines bias, while cadence controls precision.

Table 6
Exploring Time Delay Estimation Performance Against Observing Strategy

Rungs	Variable Parameter	Fixed Parameters
1, 4	Cadence (3, 6 days)	4 month seasons, 10 yr campaign
0, 3	Season (4, 8 months)	3 day cadence, 5 yr campaign
3, 4	Cadence (3,6 days)	4 month seasons, 200 epochs length
0, 1	Season (4, 8 months)	3 day cadence, 400 epochs length
1,3	Campaign (5, 10 yr)	3 day cadence, 4 month seasons
2, 3	Cadence dispersion (0, 1 days)	3 day cadence, 4 month season, 5 yr campaign

Note. The tests defined in the top two rows (above the line) are illustrated in Figure 13.

The precision of an ensemble average parameter, such as the cosmological parameters, may yet depend primarily on season length, however, through the success fraction.

These simple model conclusions represent small extrapolations—we did not, for example, test doubling the season length and cadence simultaneously—but they represent a first approximation to the response of the more accurate time delay estimation routines to variations in observing strategy.

Finally, we note briefly the implications of this model for the sample of lensed quasars that was forecast for LSST by Oguri & Marshall (2010). Rung 4 represents something like the "universal cadence" planned for LSST (Ivezic et al. 2008), albeit with slightly shorter seasons. A cadence of 6 days would be well within the reach of such a strategy, but would require using observations from most of the filters in the set. While in this work we have only simulated and analyzed single filter light curves, AGN variability has been observed to be significantly correlated across the optical and near infrared bands (see, e.g., Schmidt et al. 2012), and microlensing variability is expected, and observed, to vary smoothly with wavelength due to source size effects (e.g., Poindexter et al. 2008). With sufficiently sophisticated algorithms we might expect to be able to measure time delays from multi-filter light curves with fidelity not dissimilar to that shown by the TDC1 methods tested here. The three day cadence of Rung 1 could be achieved by LSST without changing the total number of visits; the impact of such a strategy on the various different LSST science cases would need to be investigated. We take Rungs 1 and 4 to span the range of possibilities for the LSST time sampling.

Our model suggests that, if outliers with $|A_i| > 0.1$ can be rejected (perhaps during a joint analysis of the ensemble), the cadence is effectively unimportant for time delay measurement bias, and with LSST we might expect to achieve an accuracy metric of $|A| = 0.03\%$–0.06%. Such a small time delay measurement bias is well below the systematic errors expected from lens modeling. Meanwhile, the expected precision achievable per lens in the Rung 1 and 4 cadences would be 2.6%–4.3%, and the success fractions would be 20%–26%. The mock lenses used in this data challenge were not quite randomly drawn from the OM10 catalog, but instead had approximately uniformly distributed i_3 image magnitudes within four broad magnitude bins (Section 2.2). Correcting for this, we find that we might, with the present-day algorithms (tested here and represented by our simple model), expect to be able to make time delay measurements with the above accuracy in at least 20% of an LSST sample of 1990 lenses selected to have $i_3 < 23.3$ and 10 days $< \Delta t <$ 120 days. This would correspond to a well-measured sample of around 400 lensed quasars. We must expect these numbers to be refined as the LSST observing strategy is defined, and further time delay measurement tests are carried out.

6. DISCUSSION

In this section, we give a brief analysis of each method's performance, discussing how they performed and what can be improved in the future. We note that the performance of each method must be evaluated in multi-dimensional metric space. Each "Good" Team had to make choices with respect to which metric to optimize. Some teams decided to favor inclusiveness (high *f*) at the cost of a higher fraction of outliers (lower X) or lower precision P, and vice-versa. In fact, some of the teams submitted multiple entries spanning the range of parameter space, and illustrating these competitive needs. Therefore, at this stage it is not possible, nor useful, to identify a "best" submission, not even within each method. It is more fruitful to understand the tradeoffs and explore the range covered by each method, and then identify areas for improvement.

6.1. Gaussian Processes, by Hojjati & Linder

The GP method attained its twin goals of an automated fitting pipeline and very good fit accuracy. The main issue to address is one of outliers, which can be handled in two ways: global clipping and image information. This team found that the outliers were not due to multi-modal fit distributions—indeed the fits often have better likelihood for the data than the truth. However, the cosmology derived from the outliers would be discrepant from the cosmology from the global fit ensemble, and in this way, outliers could be recognized and clipped. Another approach would be to use information such as image separation (not provided in TDC1) to recognize and discard discrepant fits. While these considerations would lower the accepted fraction of fits, the correction of the mean function discussed in Section 3.2 raises the fraction over those given here. This, and a set of new but simple selection criteria (no limits on precision were imposed by this team for TDC1 submissions), discussed in a follow-up paper by Hojjati & Linder (2014), give considerable improvement in the precision and fraction, and further improvement in accuracy.

6.2. FOT, by Romero-Wolf & Moustakas

The unblinding of the TDC1 simulated data provided valuable information on the behavior of this team's Bayesian inference algorithm. For the most part, the technique identified catastrophic outliers. However, some light curve pairs still resulted in large contributions to the χ^2 estimator. Identifying this subset of outliers that pass the quality cuts has provided valuable insight into the behavior of this technique, and will allow for future refinement and development to reduce the probability of mis-reconstructions.

6.3. Smoothing and Cross-correlation, by Aghamousa & Shafieloo

Throughout the challenge this team's main concern was to achieve a high f value without having any outliers. This was achieved with $f > 0.3$ for all five rungs. This conservative approach yielded average χ^2 values of around 0.5–0.9 for different Rungs with P of about 0.03–0.06. As noted before, since χ^2 and P are correlated, by simply dividing all estimated errors by a factor of $\sqrt{2}$, χ^2 of $\sim$1, and P of $\sim$0.02–0.04 could be achieved trivially. After the true time delays were revealed, a calibration bias of 0.5 days for all the submissions was discovered, resulting in $A \sim 1.8\%$–2.5% (the method had been calibrated only on TDC0 Rung 0, owing to lack of time). By adding a calibration correction of 0.5 days to all this team's submissions' delay estimates, the bias was removed, improving A to 0.1%–0.6%. To summarize, this method seems promising in both reliability and precision, and is automated in all steps. There is also the potential to improve the error estimation by doing appropriate simulations for each set of light curves separately.

6.4. Supervised Pelt, by Jackson

After the release of the true time delays, this submission was re-examined to try to understand the reasons for the most severe errors, especially those in which the true time delay differed from the inference by $>3\sigma$ (between 9 and 18 cases in each rung out of a few hundred submitted). In four of the worst cases, the problem appeared to be unrealistically low errors fitted during the resampling process, possibly due to a small number of anomalous points, and not corrected by eye. This suggests that for a given set of light-curves, a minimum error based on the fits to the ensemble should be adopted. A significant fraction of the remaining severe errors were characterized by a Pelt statistic versus time delay plot with a relatively bumpy and irregular minimum, even when the eye detected a good fit in terms of the number of coincident points of inflection. This is more difficult to quantify, but suggests that an addition to the resampling-derived error based on the shape of the Pelt statistic may be useful.

6.5. PyCS d3cs, spl and sdi

The `d3cs` classification of the light-curve pairs into different confidence categories proved valuable. All the resulting "doubtless" (`dou`) submissions ($f = 0.31$, averaging across all rungs) are free from any catastrophic outliers. As an example, none of the point estimates from the vanilla `spl` method is farther than $3.7\sigma_i$ or 12.0 days from the truth. For this same method, the less pure `doupla` submission ($f = 0.65$) is contaminated by 1.0% of delays that are off by more than 20 days, or, alternatively, $5\sigma_i$. Interestingly, the `d3cs` estimates for time delays shorter than 50 days are systematically biased low, leading to a significant A of approximately -0.03 for `d3cs`. We speculate that this bias is perceptual and due to users involuntarily trying to maximize the overlap in the light curves. The `sdi` and `spl` techniques were not influenced by this bias in their initial conditions, and both reached a high accuracy, consistent with being unbiased. For these two numerical techniques, the χ^2 metric values are close to unity, suggesting adequate to slightly over-estimated delay uncertainties. The implemented simplifications to the original techniques from Tewes et al. (2013a) seem therefore acceptable for the level of complexity present in the TDC1 data.

6.6. Difference-smoothing, by Rathna Kumar, Stalin, & Prabhu

From the TDC1 feedback, it was realized that this procedure overestimates the uncertainties in the measured time delays, and hence was more prone to reporting catastrophic failures. This problem can be solved by using a Gaussian filter of width equal to the median rather than the mean temporal sampling of the light curves in the process of simulating light curves having known time delays. With this choice, the intrinsic variability in the simulated light curves does not get smoothed out on short timescales. Also, there were a few cases in the submissions where the measured and true time delays were discrepant at the level of $\chi_i^2 > 10$. This points to a need to increase the plausible range of time delays around the measured delay over which the simulated light curves are generated to at least the 3σ confidence interval implied by the inferred uncertainty, rather than the 2σ confidence interval used in the TDC1 submissions. The time delay measurements can be improved further by exploring a range of reasonable values of free parameters, and selecting those which result in the smallest uncertainty in the measured time delay. These changes are now being rigorously tested on the TDC1 light curves and will be described in the paper by Rathna Kumar et al. (2014) during the revision process.

6.7. DeltaTBayes, by Tak, Meng, van Dyk, Siemiginowska, Kashyap, & Mandel

This team considered TDC1 to be a great opportunity to develop and improve their Bayesian approach. Considering the team's late entry into the challenge, the pragmatic Bayesian perspective was taken (Lee et al. 2011), developing the approximate Gibbs sampling scheme (algorithm1) and applying it only to the most realistic rung (Rung 4). The main advantage of this pragmatic approach was the fast convergence of its Markov chains, saving some computational time, a desirable characteristic for analyzing large number of data sets. The method performs well in terms of precision and accuracy. However it produces error bars that are smaller than those from a fully Bayesian approach, though larger than an empirical Bayesian approach, leading to a relatively high χ^2. To be balanced, several Gibbs sampling schemes are being tested for the future.

7. SUMMARY AND CONCLUSIONS

In the next decade, dedicated efforts and the LSST survey will deliver thousands of light curves for lensed quasars, ushering in a revolution in time-delay cosmology (Treu et al. 2013). In order to prepare for and make the most of this wealth of data, it is essential to ascertain whether current algorithms are sufficiently accurate, fast, and precise. It is also important to investigate the optimal observing strategies for time delay determination, both in dedicated monitoring campaigns and for LSST.

In order to investigate these two issues, we carried out the first strong lens TDC. After the preliminary time delay challenge TDC0 (Dobler et al. 2014), the challenge "Evil" Team simulated several thousand time delay light curves and made them available to the community on the challenge Web site. Seven "Good" Teams responded to the challenge, and blindly measured the time delays for TDC1 using 9 independent algorithms. A simple method implemented by the "Evil" Team as a baseline was also included. Our main findings from analyzing the blind TDC1 submissions can be summarized as follows.

1. The measurement of time delays from thousands of realistic light curves in manageable amounts of CPU and

investigator time has been demonstrated. This is a considerable achievement given that traditionally this process has been carried out only for very small numbers of light curves (allowing investigators to spend significant amounts of time on each system). Several independent approaches were successful, ranging from cross-correlation, to scatter minimization, to data modeling with GPs and other suitable sets of basis functions. Some methods relied heavily on visual inspection, while others were almost completely automated.

2. In Rung 0—which simulates the typical observing parameters of a dedicated monitoring campaign like COSMOGRAIL—the best current algorithms can recover time delays with negligible bias (often sub-percent) and 3% precision for over 50% of the light curves. The error bars are generally reasonable, resulting in χ^2 of the order of unity, while the fraction of outliers is also just a few percent. These were the requirements for a method to be competitive, as described in Paper I. When enough information was present in the light curves, typically six independent algorithms were able to recover time delays within 10% of the truth.
3. As the data quality was degraded in the subsequent Rungs 1–4 (emulating some observing strategies possible with LSST), the fraction of usable light curves also decreased, hovering between 20% and 30%. Outliers became more common, although they can be contained by suitably conservative algorithms, or by visual inspection. Once outliers are excluded, the algorithms perform as well as in Rung 0, albeit with a smaller fraction of the light curves (10%–30%) yielding robust results with competitive precision and accuracy. A success fraction of 20% translates to an expected sample size of around 400 lensed quasars detected and measured by LSST to very high accuracy—well within the systematic error requirements of time delay cosmography.
4. We have derived approximate scalings for the time delay metrics as a function of observing parameters. Season and campaign length appear to be the dominant terms controlling accuracy (or bias) and success rate, while the precision of the time delay is most strongly related to the cadence and campaign duration.

Much has been learned from this first blind TDC, and the results provide useful guidance and reference for designing future experiments and improving the measurement algorithms. However, it should be emphasized that this challenge was designed to be somewhat simplistic. In particular, TDC1 consisted of a pure time delay estimation challenge from light curves alone: teams were not given the image positions, nor the deflector and source redshifts. It is likely therefore that the results of this challenge might be overly pessimistic. In real life, investigators will have access to the full lensing configurations, and will be able to use this information as a prior for their time delay inference (for example using the lensing geometry for quads).

Furthermore, a fully cosmological challenge should enable outlier rejection based on cosmological self-consistency in a joint analysis of the ensemble of lenses. It should be possible to identify and reject outliers that lead to cosmological parameters (chiefly H_0) that are inconsistent with those inferred from the majority of sample. Another limitation of the simplicity of TDC1 is that the metrics measure how well an algorithm performs on time-delay estimation, not directly on cosmological parameter inference.

Given the encouraging results of TDC1, we plan to overcome these two limitations in the future. In the short term, we plan to translate the simple metrics adopted here into a full cosmological estimation tool by introducing the available additional information, and justifiable assumptions about the underlying lens models. In the longer term, we plan to organize a second TDC, to further test our ability to handle outliers, and to investigate the measurement of time delays from multi-band data, and in which more information will be provided for each system with the ultimate goal for the "Good" Teams of inferring cosmological parameters, rather than just time delays.

The TDC0 and TDC1 data will remain available at http://timedelaychallenge.org for any team who might be interested in using them for developing algorithms for strong lens time delay measurement.

We acknowledge the LSST Dark Energy Science Collaboration for hosting several meetings of the "Evil" Team, and the private code repository used in this work. We thank the referee for constructive criticism which helped improved this paper. T.T., C.D.F., and K.L. acknowledge support from the National Science Foundation collaborative grant "Collaborative Research: Accurate cosmology with strong gravitational lens time delays" (AST-1312329 and AST-1450141). T.T. gratefully acknowledges support by the Packard Foundation through a Packard Research Fellowship. K.L. is supported by China Scholarship Council. The work of P.J.M. was supported by the U.S. Department of Energy under contract number DE-AC02-76SF00515. V.B. and F.C. are supported by the Swiss National Science Foundation (SNSF). M.T. acknowledges support by the DFG grant Hi 1495/2-1. A.A and A.S. wish to acknowledge support from the Korea Ministry of Education, Science and Technology, Gyeongsangbuk-Do and Pohang City for Independent Junior Research Groups at the Asia Pacific Center for Theoretical Physics. A.S. would like to acknowledge the support of the National Research Foundation of Korea (NRF-2013R1A1A2013795). E.L. is supported by DOE grant DE-SC-0007867 and contract No. DE-AC02-05CH11231. A.H. is supported by an NSERC grant and thanks the Institute for the Early Universe, Korea, for computational resources. A.A., A.S., A.H., and E.L. thank IBS Korea for hospitality. The work of L.A.M. and A.R.W. was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. K.M. is supported at Harvard by NSF grant AST-1211196.

APPENDIX

FACTORS AFFECTING THE RMS MICROLENSING MAGNIFICATION

How sensitive is the distribution of mock light curves to the random realizations of the positions of the stars in the lens? We generated 30 star field realizations, over fields 30 Einstein Radii (R_E) by 30 R_E in area, with different random seeds for each fixed F_* or κ, and calculated the mean of their standard deviations as a characteristic measure of the rms fluctuation in the microlensing magnification. Figure 14 shows how this rms fluctuation varies as a function of F_*. The top panel shows the case where the image arises at the minimum of the time delay surface (where the eigenvalues of the Hessian matrix are both positive and the image has positive parity); the bottom panel shows the case where the image arises at a saddle point of the time delay surface (where the eigenvalues have opposite signs and the parity is flipped compared to the original source). For both figures, significant trends, increasing when F_* is small, and

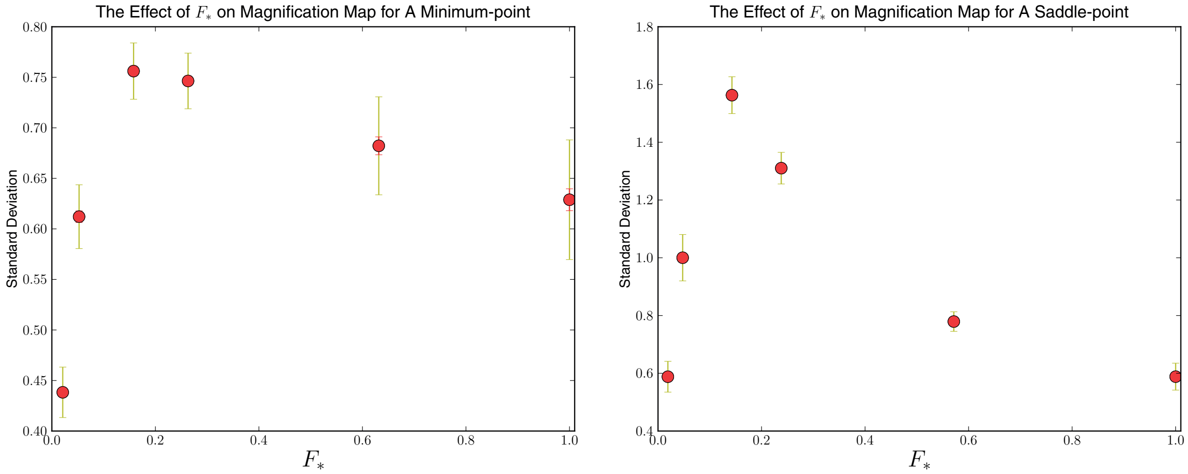


Figure 14. Mean Standard Deviation of the magnification map as a function of F_*. Each point is the result from 30 realizations with different position seeds. We show two errors: Standard Deviation (yellow) and Standard Deviation of the mean value (red). Top figure is for a minimum-image with $\kappa = 0.475$, $\gamma = 0.425$. Bottom figure is for a saddle-image with $\kappa = 0.525$, $\gamma = 0.575$. Both have the same macro magnification $\mu = 10$.

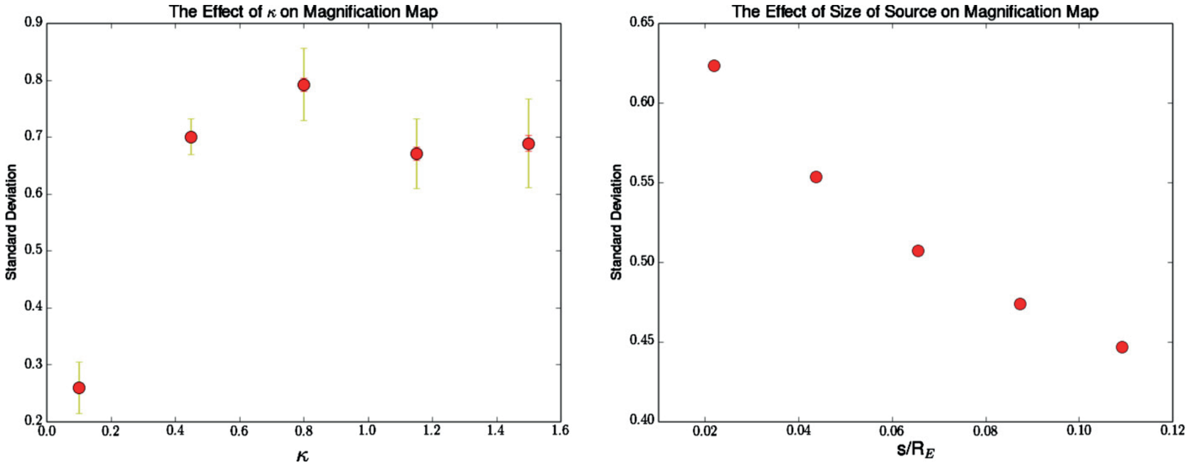


Figure 15. Mean Standard Deviation of the magnification map as a function of local convergence κ (left) and source size s (right). In the left panel, $\kappa = \gamma$ and $F_* = 0.1$ are fixed for each point, while in the right panel $\kappa = \gamma = 0.45$, $F_* = 0.1$ are fixed for each point.

decreasing at larger F_*, are apparent. These can be explained as follows.

At small F_*, when there are few stars, sparsely distributed in the field, the magnification of each position is dominated by the nearest individual star, and the variation of the map increases with more stars that bring more caustics. However, when F_* grows large, the magnification at any position becomes less affected by the addition of more stars, and the magnification and demagnification attributed to different stars will average away. The saddle-point images are more vulnerable to demagnification and hence show larger variations in their magnification maps (see Schechter 2003, for more on the differences of microlensing between minima and saddle-point images).

The left-hand panel of Figure 15 shows the effect of the macrolens convergence κ on the standard deviation of the source plane magnification map. κ affects the variation in two ways, changing the stellar density fraction, and also the macro magnification. These effects appear to approximately balance each other at high κ. At low convergence, the magnification and shear are also low, and the microlensing effects weaker.

Meanwhile, the right-hand panel of Figure 15 shows the rms microlensing magnification fluctuation as a function of source size. As expected, the fluctuations are smoothed out at large source size, reducing the amplitude of the microlensing fluctuations and ensuring that the average microlensing magnification is unity.

REFERENCES

Aghamousa, A., & Shafieloo, A. 2014, arXiv:1410.8122
Bostock, M., Ogievetsky, V., & Heer, J. 2011, IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis), 17, 2301
Coe, D., & Moustakas, L. A. 2009, ApJ, 706, 45
Dobke, B. M., King, L. J., Fassnacht, C. D., & Auger, M. W. 2009, MNRAS, 397, 311
Dobler, G., Fassnacht, C., Treu, T., et al. 2014, ApJ, in press
Fassnacht, C. D., Xanthopoulos, E., Koopmans, L. V. E., & Rusin, D. 2002, ApJ, 581, 823
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Harva, M., & Raychaudhury, S. 2006, in Proc. 16th IEEE Signal Processing Society Workshop, Machine Learning for Signal Processing, 111

Hojjati, A., Kim, A. G., & Linder, E. V. 2013, PhRvD, 87, 123512
Hojjati, A., & Linder, E. V. 2014, PhRvD, 90, 123501
Ivezic, Z., Tyson, J. A., Acosta, E., et al. 2008, arXiv:0805.2366
Kelly, B. C., Bechtold, J., & Siemiginowska, A. 2009, ApJ, 698, 895
Kochanek, C. S. 2002, ApJ, 578, 25
Lee, H., Kashyap, V. L., van Dyk, D. A., et al. 2011, ApJ, 731, 126
Linder, E. V. 2011, PhRvD, 84, 123529
LSST Dark Energy Science Collaboration 2012, arXiv:1211.0310
LSST Science Collaboration, Abell, P. A., Allison, J., Anderson, S. F., et al. 2009, arXiv:0912.0201
MacLeod, C. L., Ivezić, Ž., Kochanek, C. S., et al. 2010, ApJ, 721, 1014
Oguri, M. 2007, ApJ, 660, 1
Oguri, M., & Marshall, P. J. 2010, MNRAS, 405, 2579
Paraficz, D., & Hjorth, J. 2009, A&A, 507, L49
Paraficz, D., & Hjorth, J. 2010, ApJ, 712, 1378
Paraficz, D., Hjorth, J., Burud, I., Jakobsson, P., & Eliasdottir, A. 2006, A&A, 455, L1
Pelt, J., Hoff, W., Kayser, R., Refsdal, S., & Schramm, T. 1994, A&A, 286, 775
Planck Collaboration, Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., et al. 2014, A&A, 571, A16
Poindexter, S., Morgan, N., & Kochanek, C. S. 2008, ApJ, 673, 34
Rathna Kumar, S., Stalin, C. S., & Prabhu, T. P. 2014, submitted (arXiv:1404.2920)
Rathna Kumar, S., Tewes, M., Stalin, C. S., et al. 2013, A&A, 557, A44
Refsdal, S. 1964, MNRAS, 128, 307
Schechter, P. L. 2003, arXiv:0304480
Schechter, P. L., Wambsganss, J., & Lewis, G. F. 2004, ApJ, 613, 77
Schmidt, K. B., Rix, H.-W., Shields, J. C., et al. 2012, ApJ, 744, 147
Sereno, M., & Paraficz, D. 2014, MNRAS, 437, 600
Shafieloo, A. 2007, MNRAS, 380, 1573
Shafieloo, A. 2012, JCAP, 08, 002
Shafieloo, A., Alam, U., Sahni, V., & Starobinsky, A. A. 2006, MNRAS, 366, 1081
Shafieloo, A., & Clarkson, C. 2010, PhRvD, 81, 083537
Stoughton, C., Lupton, R. H., Bernardi, M., et al. 2002, AJ, 123, 485
Suyu, S. H., Auger, M. W., Hilbert, S., et al. 2013, ApJ, 766, 70
Suyu, S. H., Treu, T., Blandford, R. D., et al. 2012, arXiv:1202.4459
Suyu, S. H., Treu, T., Hilbert, S., et al. 2014, ApJL, 788, L35
Tewes, M., Courbin, F., & Meylan, G. 2013a, A&A, 553, A120
Tewes, M., Courbin, F., Meylan, G., et al. 2013b, A&A, 556, A22
Treu, T. 2010, ARA&A, 48, 87
Treu, T., Marshall, P. J., Cyr-Racine, F.-Y., et al. 2013, arXiv:1306.1272
Weinberg, D. H., Mortonson, M. J., Eisenstein, D. J., et al. 2013, PhR, 530, 87

B H0LiCOW Press Release

Man is profoundly dependent on the reflection of himself in another man's soul,
be it even the soul of an idiot.

- Witold Gombrowitcz, Ferdydurke -

I already mentioned in the main part of this thesis the chance I had to be a member of the H0LiCOW collaboration, at the time such nice results were ready for publication. I must say I am even more happy about the decision to prepare a press release about it, and to have pushed for adding more fanciness to it. Indeed, in addition to the usual few paragraphs of text and pretty pictures going along, we filmed a ~ 8 minutes explanatory video, making use of the power of the MOOC studio at EPFL[1] and the expertise of all the people working there. The resulting clip has been largely diffused along with the press release, accumulating dozens of thousands of views. A link to its Youtube version is given in the table that follows.

The press release aired in January 2017 and appeared on all the major professional scientific institutions' website (ESA, NASA, etc...), alongside all the home institutions of all the scientists involved in H0LiCOW. The general media - mainstream newspapers, astronomy magazines, tech and science blogs and even radio news - were also keen on diffusing the press release as well as, in some cases, build a story on it and get in touch with us to add a more personal taste to it. For a while, I tried to keep a record of the links of all the articles and interviews mentioning H0LiCOW, but at some point it grew so big that I had to resign and simply enjoy the fact that news about our work was spreading in the whole world. Before stopping, I counted 60+ original articles and 100+ simple copy/paste of the press release or of one of the original articles, principally written in English or French. I regrouped in Tab. B.1 web links to some of the most prestigious media articles about our work.

Working on the press release and the video took me a significant amount of time, yet I grew a lot in the process. The exercise of simplifying my research to the point anyone without any specific background can understand it, while distilling enough accurate information to

---

[1] https://moocs.epfl.ch/mooc-factory

keep the experts on the field interested is much more tricky than it seems in first approach. In addition, confronting my way of thinking and interpreting the scientific results to the outbursts of sensationalism of some of the journalists making a living from it has been a very interesting experiment. To say the least, it makes you read everything you do not have a specific knowledge on with a suspicious eye. But in general, working on that press release has been a very positive experience that I would without any doubt reiterate in the future.

Table B.1: Selection of English and French press articles relating the H0LiCOW news about the measurement of the Hubble constant presented in Bonvin et al. (2017). To access the article, either click on the journal name, or replace *** by the corresponding keyword in the url http://cpc.cx/*** (i.e. http://cpc.cx/jK4) for the ESA's press release article.

Media	Headline	Keyword
ESA	Cosmic lenses support finding on faster than expected expansion of the Universe	jK4
NASA	Cosmic lenses support finding on faster than expected expansion of the Universe	jK5
Royal Astronomical Society	Cosmic lenses support finding on faster than expected expansion of the Universe	jK6
Youtube	H0LiCOW - H0 Lenses in COsmograil's Wellspring	jKr
EPFL	How fast is the universe expanding? Quasars provide an answer	jK8
The Guardian	Speedy universe expansion challenges Einstein's theory	jK9
New York Times	Cosmos Controversy: The Universe Is Expanding, but How Fast?	jKp
Le Figaro (in French)	L'Univers s'étendrait plus vite que prévu	jKe
Le Temps (in French)	Notre Univers s'étend plus vite que prévu, par la faute de l'energie noire	jKd
The Daily Mail	The universe is expanding FASTER than expected - and astronomers say something 'beyond our current knowledge' is causing it	jKf
International Business Times	Universe expansion rate creates cosmological problem beyond our current understanding	jKg
Huffington Post	The Universe Is Expanding So Fast It Could Rip Galaxies Apart, New Research Suggests	jKm
Phys.org	Astronomers measure universe expansion, get hints of 'new physics'	jKq
Gizmodo	New Measurements of the Universe Expanding Tell a Confusing Story	jKa
Space	The Universe Is Expanding Surprisingly Fast	jKb
Inverse	A New Measure of the Expanding Universe May Lead to a 'New Physics'	jKc
Vice Motherboard	The Expansion of the Universe Appears Faster Than We Thought	jKn
Sky & Telescope	H0Li COW! Is Our Universe Expanding Faster Than We Thought?	jKo

Bibliography

Addison, G.E., Huang, Y., Watts, D.J., Bennett, C.L., Halpern, M., Hinshaw, G. and Weiland, J.L. (2016). Quantifying Discordance in the 2015 Planck CMB Spectrum. *ApJ*, 818:132.

Aghamousa, A. and Shafieloo, A. (2015). Fast and Reliable Time Delay Estimation of Strong Lens Systems Using the Smoothing and Cross-correlation Methods. *ApJ*, 804:39.

Agnello, A., Kelly, B.C., Treu, T. and Marshall, P.J. (2015). Data mining for gravitationally lensed quasars. *MNRAS*, 448:1446–1462.

Agnello, A., Lin, H., Buckley-Geer, L., Treu, T., Bonvin, V. et al (2017). Models of the strongly lensed quasar DES J0408-5354. *ArXiv e-prints:1702.00406*.

Alam, S., Ata, M., Bailey, S., Beutler, F., Bizyaev, D. et al (2016). The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample. *ArXiv e-prints:1607.03155*.

Alcock, C., Akerlof, C.W., Allsman, R.A., Axelrod, T.S., Bennett, D.P. et al (1993). Possible gravitational microlensing of a star in the Large Magellanic Cloud. *Nature*, 365:621–623.

Alpher, R.A. and Herman, R. (1948). Evolution of the Universe. *Nature*, 162:774–775.

Astier, P., Guy, J., Regnault, N., Pain, R., Aubourg, E. et al (2006). The Supernova Legacy Survey: measurement of Ω_M, Ω_Λ and w from the first year data set. *A&A*, 447:31–48.

Astropy Collaboration, Robitaille, T.P., Tollerud, E.J., Greenfield, P., Droettboom, M. et al (2013). Astropy: A community Python package for astronomy. *A&A*, 558:A33.

Aubourg, E., Bareyre, P., Bréhin, S., Gros, M., Lachièze-Rey, M. et al (1993). Evidence for gravitational microlensing by dark objects in the Galactic halo. *Nature*, 365:623–625.

Avestruz, C., Li, N., Lightman, M., Collett, T.E. and Luo, W. (2017). Automated Lensing Learner - I: An Automated Strong Lensing Identification Pipeline. *ArXiv e-prints:1704.02322*.

Bacon, D.J., Refregier, A.R. and Ellis, R.S. (2000). Detection of weak gravitational lensing by large-scale structure. *MNRAS*, 318:625–640.

Barkana, R. (1997). Analysis of Time Delays in the Gravitational Lens PG 1115+080. *ApJ*, 489:21–28.

Beaton, R.L., Freedman, W.L., Madore, B.F., Bono, G., Carlson, E.K. et al (2016). The Carnegie-Chicago Hubble Program. I. An Independent Approach to the Extragalactic Distance Scale Using Only Population II Distance Indicators. *ApJ*, 832:210.

Beckers, J.M. (1993). Adaptive optics for astronomy - Principles, performance, and applications. *ARA&A*, 31:13–62.

Bennett, C.L., Banday, A.J., Gorski, K.M., Hinshaw, G., Jackson, P. et al (1996). Four-Year COBE DMR Cosmic Microwave Background Observations: Maps and Basic Results. *ApJ*, 464:L1.

Bennett, C.L., Bay, M., Halpern, M., Hinshaw, G., Jackson, C. et al (2003). The Microwave Anisotropy Probe Mission. *ApJ*, 583:1–23.

Bennett, C.L., Larson, D., Weiland, J.L. and Hinshaw, G. (2014). The 1% Concordance Hubble Constant. *ApJ*, 794:135.

Bennett, C.L., Larson, D., Weiland, J.L., Jarosik, N., Hinshaw, G. et al (2013a). Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results. *ApJS*, 208:20.

Bennett, C.L., Larson, D., Weiland, J.L., Jarosik, N., Hinshaw, G. et al (2013b). Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results. *ApJS*, 208:20.

Bernal, J.L., Verde, L. and Riess, A.G. (2016). The trouble with H_0. *J. Cosmology Astropart. Phys.*, 10:019.

Bernstein, G. and Fischer, P. (1999). Values of H_0 from Models of the Gravitational Lens 0957+561. *AJ*, 118:14–34.

Bertin, E. and Arnouts, S. (1996). SExtractor: Software for source extraction. *A&AS*, 117:393–404.

Bethapudi, S. and Desai, S. (2017). Median statistics estimates of Hubble and Newton's Constant. *ArXiv e-prints:1701.01789*.

Betoule, M., Kessler, R., Guy, J., Mosher, J., Hardin, D. et al (2014). Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples. *A&A*, 568:A22.

Beutler, F., Blake, C., Colless, M., Jones, D.H., Staveley-Smith, L. et al (2011). The 6dF Galaxy Survey: baryon acoustic oscillations and the local Hubble constant. *MNRAS*, 416:3017–3032.

Biggs, A.D., Browne, I.W.A., Helbig, P., Koopmans, L.V.E., Wilkinson, P.N. and Perley, R.A. (1999). Time delay for the gravitational lens system B0218+357. *MNRAS*, 304:349–358.

Binney, J. and Tremaine, S. (1987). *Galactic dynamics*.

Birrer, S., Amara, A. and Refregier, A. (2016). The mass-sheet degeneracy and time-delay cosmography: analysis of the strong lens RXJ1131-1231. *J. Cosmology Astropart. Phys.*, 8:020.

Blake, C., Kazin, E.A., Beutler, F., Davis, T.M., Parkinson, D. et al (2011). The WiggleZ Dark Energy Survey: mapping the distance-redshift relation with baryon acoustic oscillations. *MNRAS*, 418:1707–1724.

Blandford, R.D. and Narayan, R. (1992). Cosmological applications of gravitational lensing. *ARA&A*, 30:311–358.

Boffin, H., Blanchard, G., Gonzalez, O., Moehler, S., Sedaghati, E. et al (2015). Making FORS2 Fit for Exoplanet Observations (again). *The Messenger*, 159:6–9.

Boffin, H., Moehler, S. and Freudling, W. (2016). FORS2 Rotating Flat Field Systematics Fixed - Recent Exchange of FORS LADC Prisms Improves the Long-known Flat-fielding Problem. *The Messenger*, 163:10–11.

Bonamente, M., Joy, M.K., LaRoque, S.J., Carlstrom, J.E., Reese, E.D. and Dawson, K.S. (2006). Determination of the Cosmic Distance Scale from Sunyaev-Zel'dovich Effect and Chandra X-Ray Measurements of High-Redshift Galaxy Clusters. *ApJ*, 647:25–54.

Bond, J.R. and Efstathiou, G. (1984). Cosmic background radiation anisotropies in universes dominated by nonbaryonic dark matter. *ApJ*, 285:L45–L48.

Bondi, H. and Gold, T. (1948). The Steady-State Theory of the Expanding Universe. *MNRAS*, 108:252.

Bono, G. (2003). RR Lyrae Distance Scale: Theory and Observations. In Alloin, D. and Gieren, W., editors, *Stellar Candles for the Extragalactic Distance Scale*, volume 635 of *Lecture Notes in Physics, Berlin Springer Verlag*, pages 85–104.

Bonvin, V., Courbin, F., Suyu, S.H., Marshall, P.J., Rusu, C.E. et al (2017). H0LiCOW - V. New COSMOGRAIL time delays of HE 0435-1223: H_0 to 3.8 per cent precision from strong lensing in a flat ΛCDM model. *MNRAS*, 465:4914–4930.

Bonvin, V., Tewes, M., Courbin, F., Kuntzer, T., Sluse, D. and Meylan, G. (2016). COSMOGRAIL: the COSmological MOnitoring of GRAvItational Lenses. XV. Assessing the achievability and precision of time-delay measurements. *A&A*, 585:A88.

Braatz, J.A., Reid, M.J., Humphreys, E.M.L., Henkel, C., Condon, J.J. and Lo, K.Y. (2010). The Megamaser Cosmology Project. II. The Angular-diameter Distance to UGC 3789. *ApJ*, 718:657–665.

Burke, W.L. (1981). Multiple Gravitational Imaging by Distributed Masses. *ApJ*, 244:L1.

Burud, I., Courbin, F., Magain, P., Lidman, C., Hutsemékers, D. et al (2002). An optical time-delay for the lensed BAL quasar HE 2149-2745. *A&A*, 383:71–81.

Calabrese, E., Hlozek, R.A., Battaglia, N., Battistelli, E.S., Bond, J.R. et al (2017). Cosmological parameters from pre-planck cmb measurements: A 2017 update. *Phys. Rev. D*, 95:063525.

Cantale, N., Courbin, F., Tewes, M., Jablonka, P. and Meylan, G. (2016). Firedec: a two-channel finite-resolution image deconvolution algorithm. *A&A*, 589:A81.

Cardona, W., Kunz, M. and Pettorino, V. (2017). Determining H_0 with Bayesian hyper-parameters. *J. Cosmology Astropart. Phys.*, 3:056.

Carlstrom, J.E., Ade, P.A.R., Aird, K.A., Benson, B.A., Bleem, L.E. et al (2011). The 10 Meter South Pole Telescope. *PASP*, 123:568.

Cavaliere, A., Danese, L. and de Zotti, G. (1979). Cosmic distances from X-ray and microwave observations of clusters of galaxies. *A&A*, 75:322–325.

Chan, J.H.H., Suyu, S.H., Chiueh, T., More, A., Marshall, P.J., Coupon, J., Oguri, M. and Price, P. (2015). Chitah: Strong-gravitational-lens Hunter in Imaging Surveys. *ApJ*, 807:138.

Chan, J.H.H., Suyu, S.H., More, A., Oguri, M., Chiueh, T. et al (2016). Galaxy-scale Gravitational Lens Candidates from the Hyper Suprime-Cam Imaging Survey and the Galaxy And Mass Assembly Spectroscopic Survey. *ApJ*, 832:135.

Chang, K. and Refsdal, S. (1979). Flux variations of QSO 0957+561 A, B and image splitting by stars near the light path. *Nature*, 282:561–564.

Chantry, V., Sluse, D. and Magain, P. (2010). COSMOGRAIL: the COSmological MOnitoring of GRAvItational Lenses. VIII. Deconvolution of high resolution near-IR images and simple mass models for 7 gravitationally lensed quasars. *A&A*, 522:A95.

Chartas, G., Brandt, W.N. and Gallagher, S.C. (2003). XMM-Newton Reveals the Quasar Outflow in PG 1115+080. *ApJ*, 595:85–93.

Chartas, G., Brandt, W.N., Gallagher, S.C. and Proga, D. (2007). XMM-Newton and Chandra Spectroscopy of the Variable High-Energy Absorption of PG 1115+080: Refined Outflow Constraints. *AJ*, 133:1849–1860.

Chartas, G., Kochanek, C.S., Dai, X., Poindexter, S. and Garmire, G. (2009). X-Ray Microlensing in RXJ1131-1231 and HE1104-1805. *ApJ*, 693:174–185.

Chávez, R., Terlevich, E., Terlevich, R., Plionis, M., Bresolin, F., Basilakos, S. and Melnick, J. (2012). Determining the Hubble constant using giant extragalactic H II regions and H II galaxies. *MNRAS*, 425:L56–L60.

Coccato, L., Bramich, D.M., Freudling, W. and Moehler, S. (2014). Removal of systematics in photometric measurements: static and rotating illumination corrections in FORS2@VLT data. *MNRAS*, 438:1256–1266.

Coles, J. (2008). A New Estimate of the Hubble Time with Improved Modeling of Gravitational Lenses. *ApJ*, 679:17–24.

Collett, T.E. (2015). The Population of Galaxy-Galaxy Strong Lenses in Forthcoming Optical Imaging Surveys. *ApJ*, 811:20.

Collett, T.E. and Cunnington, S.D. (2016). Observational selection biases in time-delay strong lensing and their impact on cosmography. *MNRAS*, 462:3255–3264.

Collett, T.E., Marshall, P.J., Auger, M.W., Hilbert, S., Suyu, S.H. et al (2013). Reconstructing the lensing mass in the Universe from photometric catalogue data. *MNRAS*, 432:679–692.

Copi, C.J., Huterer, D., Schwarz, D.J. and Starkman, G.D. (2006). On the large-angle anomalies of the microwave sky. *MNRAS*, 367:79–102.

Copi, C.J., Huterer, D., Schwarz, D.J. and Starkman, G.D. (2015). Lack of large-angle TT correlations persists in WMAP and Planck. *MNRAS*, 451:2978–2985.

Courbin, F. and Bonvin, V. (in prep.). eCOSMOGRAIL : extended COSMOGRAIL, speeding-up the measurement of time-delays.

Courbin, F., Bonvin, V., Buckley-Geer, E., Fassnacht, C.D., Frieman, J. et al (2017). COSMOGRAIL XVI: Time delays for the quadruply imaged quasar DES J0408-5354 with high-cadence photometric monitoring. *ArXiv e-prints:1706.09424*.

Courbin, F., Chantry, V., Revaz, Y., Sluse, D., Faure, C. et al (2011). COSMOGRAIL: the COSmological MOnitoring of GRAvItational Lenses. IX. Time delays, lens dynamics and baryonic fraction in HE 0435-1223. *A&A*, 536:A53.

Courbin, F., Lidman, C. and Magain, P. (1998). Detection of the lensing galaxy in HE 1104-1805. *A&A*, 330:57–62.

Courbin, F., Magain, P., Keeton, C.R., Kochanek, C.S., Vanderriest, C., Jaunsen, A.O. and Hjorth, J. (1997). The geometry of the quadruply imaged quasar PG 1115+080: implications for H_0_. *A&A*, 324:L1–L4.

Courtois, H.M. and Tully, R.B. (2012). Cosmicflows-2: Type Ia Supernova Calibration and H $_0$. *ApJ*, 749:174.

Cox, J.P. (1980). *Theory of stellar pulsation*.

Croft, R.A.C. and Dailey, M. (2011). On the measurement of cosmological parameters. *ArXiv e-prints:1112.3108*.

da Cunha, E., Hopkins, A.M., Colless, M., Taylor, E.N., Blake, C. et al (2017). The Taipan Galaxy Survey: Scientific Goals and Observing Strategy. *ArXiv e-prints:1706.01246*.

Dai, X., Kochanek, C.S., Chartas, G., Kozłowski, S., Morgan, C.W., Garmire, G. and Agol, E. (2010). The Sizes of the X-ray and Optical Emission Regions of RXJ 1131-1231. *ApJ*, 709:278–285.

Dalal, N. and Kochanek, C.S. (2002). Direct Detection of Cold Dark Matter Substructure. *ApJ*, 572:25–33.

Dark Energy Survey Collaboration, Abbott, T., Abdalla, F.B., Aleksić, J., Allam, S. et al (2016). The Dark Energy Survey: more than dark energy - an overview. *MNRAS*, 460:1270–1299.

de Grijs, R., Courbin, F., Martinez-Vazquez, C.E., Monelli, M., Oguri, M. and Suyu, S.H. (2017). Toward an internally consistent astronomical distance scale. *ArXiv e-prints:1706.07933*.

de Jong, J.T.A., Verdoes Kleijn, G.A., Kuijken, K.H. and Valentijn, E.A. (2013). The Kilo-Degree Survey. *Experimental Astronomy*, 35:25–44.

de Oliveira-Costa, A. and Tegmark, M. (2006). CMB multipole measurements in the presence of foregrounds. *Phys. Rev. D*, 74(2):023005.

de Vaucouleurs, G. and Bollinger, G. (1979). The extragalactic distance scale. VII - The velocity-distance relations in different directions and the Hubble ratio within and without the local supercluster. *ApJ*, 233:433–452.

DES Collaboration, Abbott, T.M.C., Abdalla, F.B., Alarcon, A., Aleksić, J. et al (2017). Dark Energy Survey Year 1 Results: Cosmological Constraints from Galaxy Clustering and Weak Lensing. *ArXiv e-prints:1708.01530*.

Dewdney, P.E., Hall, P.J., Schilizzi, R.T. and Lazio, T.J.L.W. (2009). The Square Kilometre Array. *IEEE Proceedings*, 97:1482–1496.

Dhawan, S., Jha, S.W. and Leibundgut, B. (2017). Measuring the Hubble constant with Type Ia supernovae as near-infrared standard candles. *ArXiv e-prints:1707.00715*.

Di Valentino, E., Melchiorri, A., Linder, E.V. and Silk, J. (2017). Constraining Dark Energy Dynamics in Extended Parameter Space. *ArXiv e-prints:1704.00762*.

Di Valentino, E., Melchiorri, A. and Silk, J. (2016). Reconciling Planck with the local value of H_0 in extended parameter space. *Physics Letters B*, 761:242–246.

Diego, J.M., Broadhurst, T., Chen, C., Lim, J., Zitrin, A. et al (2016). A free-form prediction for the reappearance of supernova Refsdal in the Hubble Frontier Fields cluster MACSJ1149.5+2223. *MNRAS*, 456:356–365.

Diehl, H.T. (2017). The DES Bright Arcs Survey: Hundreds of Candidate Strongly Lensed Galaxy Systems from the Dark Energy Survey Science Verification and Year 1 Observations. *Submitted to: Astrophys. J. Suppl.*

Ding, X., Liao, K., Treu, T., Suyu, S.H., Chen, G.C.F. et al (2017a). H0LiCOW. VI. Testing the fidelity of lensed quasar host galaxy reconstruction. *MNRAS*, 465:4634–4649.

Ding, X., Treu, T., Suyu, S.H., Wong, K.C., Morishita, T. et al (2017b). H0LiCOW VII. Cosmic evolution of the correlation between black hole mass and host galaxy luminosity. *ArXiv e-prints:1703.02041*.

Dobler, G., Fassnacht, C.D., Treu, T., Marshall, P., Liao, K., Hojjati, A., Linder, E. and Rumbaugh, N. (2015). Strong Lens Time Delay Challenge. I. Experimental Design. *ApJ*, 799:168.

Doux, C., Penna-Lima, M., Vitenti, S.D.P., Tréguer, J., Aubourg, E. and Ganga, K. (2017). Cosmological constraints from a joint analysis of cosmic microwave background and large-scale structure. *ArXiv e-prints:1706.04583*.

Dyson, F.W., Eddington, A.S. and Davidson, C. (1920). A Determination of the Deflection of Light by the Sun's Gravitational Field, from Observations Made at the Total Eclipse of May 29, 1919. *Philosophical Transactions of the Royal Society of London Series A*, 220:291–333.

Eddington, A.S. (1930). On the instability of Einstein's spherical world. *MNRAS*, 90:668–678.

Efstathiou, G. (2014). H_0 revisited. *MNRAS*, 440:1138–1152.

Efstathiou, G. and Lemos, P. (2017). Problems with KiDS. *ArXiv e-prints:1707.00483*.

Efstathiou, G., Sutherland, W.J. and Maddox, S.J. (1990). The cosmological constant and cold dark matter. *Nature*, 348:705–707.

Eigenbrod, A., Courbin, F., Dye, S., Meylan, G., Sluse, D., Vuissoz, C. and Magain, P. (2006a). COSMOGRAIL: the COSmological MOnitoring of GRAvItational Lenses. II. SDSS J0924+0219: the redshift of the lensing galaxy, the quasar spectral variability and the Einstein rings. *A&A*, 451:747–757.

Eigenbrod, A., Courbin, F. and Meylan, G. (2007). COSMOGRAIL: the COSmological MOnitoring of GRAvItational Lenses. VI. Redshift of the lensing galaxy in seven gravitationally lensed quasars. *A&A*, 465:51–56.

Eigenbrod, A., Courbin, F., Meylan, G., Agol, E., Anguita, T., Schmidt, R.W. and Wambsganss, J. (2008). Microlensing variability in the gravitationally lensed quasar QSO 2237+0305 - the Einstein Cross. II. Energy profile of the accretion disk. *A&A*, 490:933–943.

Eigenbrod, A., Courbin, F., Meylan, G., Vuissoz, C. and Magain, P. (2006b). COSMOGRAIL: the COSmological MOnitoring of GRAvItational Lenses. III. Redshift of the lensing galaxy in eight gravitationally lensed quasars. *A&A*, 451:759–766.

Eigenbrod, A., Courbin, F., Vuissoz, C., Meylan, G., Saha, P. and Dye, S. (2005). COSMOGRAIL: The COSmological MOnitoring of GRAvItational Lenses. I. How to sample the light curves of gravitationally lensed quasars to measure accurate time delays. *A&A*, 436:25–35.

Einstein, A. (1905). Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Annalen der Physik*, 322:132–148.

Einstein, A. (1915a). Erklarung der Perihelionbewegung der Merkur aus der allgemeinen Relativitatstheorie. *Sitzungsber. preuss.Akad. Wiss., vol. 47, No.2, pp. 831-839, 1915*, 47:831–839.

Einstein, A. (1915b). Zur allgemeinen Relativitätstheorie. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 778-786.*

Einstein, A. (1917). Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 142-152.*

Einstein, A. (1936). Lens-like action of a star by the deviation of light in the gravitational field. *Science*, 84(2188):506–507.

Eisenstein, D.J., Zehavi, I., Hogg, D.W., Scoccimarro, R., Blanton, M.R. et al (2005). Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. *ApJ*, 633:560–574.

Eulaers, E. and Magain, P. (2011). Time delays for eleven gravitationally lensed quasars revisited. *A&A*, 536:A44.

Eulaers, E., Tewes, M., Magain, P., Courbin, F., Asfandiyarov, I. et al (2013). COSMOGRAIL: the COSmological MOnitoring of GRAvItational Lenses. XII. Time delays of the doubly lensed quasars SDSS J1206+4332 and HS 2209+1914. *A&A*, 553:A121.

Fadely, R., Keeton, C.R., Nakajima, R. and Bernstein, G.M. (2010). Improved Constraints on the Gravitational Lens Q0957+561. II. Strong Lensing. *ApJ*, 711:246–267.

Falco, E.E., Gorenstein, M.V. and Shapiro, I.I. (1985). On model-dependent bounds on H(0) from gravitational images Application of Q0957 + 561A,B. *ApJ*, 289:L1–L4.

Falco, E.E., Shapiro, I.I., Moustakas, L.A. and Davis, M. (1997). An Estimate of H_0 from Keck Spectroscopy of the Gravitational Lens System 0957+561. *ApJ*, 484:70–78.

Falco, E.E., Wambsganss, J. and Schneider, P. (1991). The role of microlensing in estimates of the relative time delay for the gravitational images of Q0957 + 561. *MNRAS*, 251:698–706.

Fassnacht, C.D., Gal, R.R., Lubin, L.M., McKean, J.P., Squires, G.K. and Readhead, A.C.S. (2006). Mass along the Line of Sight to the Gravitational Lens B1608+656: Galaxy Groups and Implications for H_0. *ApJ*, 642:30–38.

Fassnacht, C.D., Koopmans, L.V.E. and Wong, K.C. (2011). Galaxy number counts and implications for strong lensing. *MNRAS*, 410:2167–2179.

Fassnacht, C.D., Xanthopoulos, E., Koopmans, L.V.E. and Rusin, D. (2002). A Determination of H_0 with the CLASS Gravitational Lens B1608+656. III. A Significant Improvement in the Precision of the Time Delay Measurements. *ApJ*, 581:823–835.

Feeney, S.M., Mortlock, D.J. and Dalmasso, N. (2017). Clarifying the Hubble constant tension with a Bayesian hierarchical model of the local distance ladder. *ArXiv e-prints:1707.00007*.

Fohlmeister, J., Kochanek, C.S., Falco, E.E., Morgan, C.W. and Wambsganss, J. (2008). The Rewards of Patience: An 822 Day Time Delay in the Gravitational Lens SDSS J1004+4112. *ApJ*, 676:761–766.

Foreman-Mackey, D., Agol, E., Angus, R. and Ambikasaran, S. (2017). Fast and scalable Gaussian process modeling with applications to astronomical time series. *ArXiv e-prints:1703.09710*.

Fowler, J.W., Acquaviva, V., Ade, P.A.R., Aguirre, P., Amiri, M. et al (2010). The Atacama Cosmology Telescope: A Measurement of the 600 ell 8000 Cosmic Microwave Background Power Spectrum at 148 GHz. *ApJ*, 722:1148–1161.

Freedman, W.L. (2017). Cosmology at a crossroads. *Nature Astronomy*, 1:0121.

Freedman, W.L., Madore, B.F., Gibson, B.K., Ferrarese, L., Kelson, D.D. et al (2001). Final Results from the Hubble Space Telescope Key Project to Measure the Hubble Constant. *ApJ*, 553:47–72.

Freedman, W.L., Madore, B.F., Scowcroft, V., Burns, C., Monson, A., Persson, S.E., Seibert, M. and Rigby, J. (2012). Carnegie Hubble Program: A Mid-infrared Calibration of the Hubble Constant. *ApJ*, 758:24.

Friedmann, A. (1922). Über die Krümmung des Raumes. *Zeitschrift fur Physik*, 10:377–386.

Frieman, J.A., Bassett, B., Becker, A., Choi, C., Cinabro, D. et al (2008). The Sloan Digital Sky Survey-II Supernova Survey: Technical Summary. *AJ*, 135:338–347.

Galama, T.J., Vreeswijk, P.M., van Paradijs, J., Kouveliotou, C., Augusteijn, T. et al (1998). An unusual supernova in the error box of the γ-ray burst of 25 April 1998. *Nature*, 395:670–672.

Gamow, G. (1948). The Evolution of the Universe. *Nature*, 162:680–682.

Gao, F., Braatz, J.A., Reid, M.J., Lo, K.Y., Condon, J.J. et al (2016). The Megamaser Cosmology Project. VIII. A Geometric Distance to NGC 5765b. *ApJ*, 817:128.

Giannini, E., Schmidt, R.W., Wambsganss, J., Alsubai, K., Andersen, J.M. et al (2017). MiNDSTEp differential photometry of the gravitationally lensed quasars WFI 2033-4723 and HE 0047-1756: microlensing and a new time delay. *A&A*, 597:A49.

Gil-Merino, R., Wisotzki, L. and Wambsganss, J. (2002). The Double Quasar HE 1104-1805: A case study for time delay determination with poorly sampled lightcurves. *A&A*, 381:428–439.

Goldstein, D.A., Nugent, P.E., Kasen, D.N. and Collett, T.E. (2017). Precise Time Delays from Chromatically Microlensed Type Ia Supernovae. *ArXiv e-prints:1708.00003.*

Goobar, A., Amanullah, R., Kulkarni, S.R., Nugent, P.E., Johansson, J. et al (2017). iPTF16geu: A multiply imaged, gravitationally lensed type Ia supernova. *Science*, 356:291–295.

Gorenstein, M.V., Shapiro, I.I. and Falco, E.E. (1988). Degeneracies in parameter estimates for models of gravitational lens systems. *ApJ*, 327:693–711.

Gott, III, J.R. (1981). Are heavy halos made of low mass stars - A gravitational lens test. *ApJ*, 243:140–146.

Gould, A. and Loeb, A. (1992). Discovering planetary systems through gravitational microlenses. *ApJ*, 396:104–114.

Greene, Z.S., Suyu, S.H., Treu, T., Hilbert, S., Auger, M.W. et al (2013). Improving the Precision of Time-delay Cosmography with Observations of Galaxies along the Line of Sight. *ApJ*, 768:39.

Grogin, N.A. and Narayan, R. (1996). A New Model of the Gravitational Lens 0957+561 and a Limit on the Hubble Constant. *ApJ*, 464:92.

Guy, J., Sullivan, M., Conley, A., Regnault, N., Astier, P. et al (2010). The Supernova Legacy Survey 3-year sample: Type Ia supernovae photometric distances and cosmological constraints. *A&A*, 523:A7.

Harrison, E.R. (1970). Fluctuations at the Threshold of Classical Cosmology. *Phys. Rev. D*, 1:2726–2730.

Harvey, D., Courbin, F., Kneib, J.P. and McCarthy, I.G. (2017). A detection of wobbling Brightest Cluster Galaxies within massive galaxy clusters. *ArXiv e-prints:1703.07365.*

Hatt, D., Beaton, R.L., Freedman, W.L., Madore, B.F., Jang, I.S. et al (2017). The Carnegie-Chicago Hubble Program. II. The Distance to IC 1613: The Tip of the Red Giant Branch and RR Lyrae Period-Luminosity Relations. *ArXiv e-prints:1703.06468.*

Hazard, C., Mackey, M.B. and Shimmins, A.J. (1963). Investigation of the Radio Source 3C 273 By The Method of Lunar Occultations. *Nature*, 197:1037–1039.

Heavens, A., Fantaye, Y., Sellentin, E., Eggers, H., Hosenie, Z., Kroon, S. and Mootoovaloo, A. (2017). No evidence for extensions to the standard cosmological model. *ArXiv e-prints:1704.03467.*

Hege, E.K., Hubbard, E.N., Strittmatter, P.A. and Worden, S.P. (1981). Speckle interferometry observations of the triple QSO PG 1115 + 08. *ApJ*, 248:L1–L3.

Henry, J.P. and Heasley, J.N. (1986). High-resolution imaging from Mauna Kea - The triple quasar in 0.3-arc S seeing. *Nature*, 321:139–142.

Herrnstein, J.R., Moran, J.M., Greenhill, L.J., Diamond, P.J., Inoue, M. et al (1999). A geometric distance to the galaxy NGC4258 from orbital motions in a nuclear gas disk. *Nature*, 400:539–541.

Hilbert, S., Hartlap, J., White, S.D.M. and Schneider, P. (2009). Ray-tracing through the Millennium Simulation: Born corrections and lens-lens coupling in cosmic shear and galaxy-galaxy lensing. *A&A*, 499:31–43.

Hilbert, S., White, S.D.M., Hartlap, J. and Schneider, P. (2007). Strong lensing optical depths in a ΛCDM universe. *MNRAS*, 382:121–132.

Hildebrandt, H., Viola, M., Heymans, C., Joudaki, S., Kuijken, K. et al (2017). KiDS-450: cosmological parameter constraints from tomographic weak gravitational lensing. *MNRAS*, 465:1454–1498.

Hinshaw, G., Larson, D., Komatsu, E., Spergel, D.N., Bennett, C.L. et al (2013). Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results. *ApJS*, 208:19.

Hirv, A., Olspert, N. and Pelt, J. (2011). Towards the Automatic Estimation of Time Delays of Gravitational Lenses. *Baltic Astronomy*, 20:125–144.

Hjorth, J., Burud, I., Jaunsen, A.O., Schechter, P.L., Kneib, J.P. et al (2002). The Time Delay of the Quadruple Quasar RX J0911.4+0551. *ApJ*, 572:L11–L14.

Hojjati, A. and Linder, E.V. (2014). Next generation strong lensing time delay estimation with Gaussian processes. *Phys. Rev. D*, 90(12):123501.

Hook, I.M., McMahon, R.G., Boyle, B.J. and Irwin, M.J. (1994). The Variability of Optically Selected Quasars. *MNRAS*, 268:305.

Hoyle, F. and Sandage, A. (1956). The Second-Order Term in the Redshift-Magnitude Relation. *PASP*, 68:301.

Hubble, E. (1929). A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15:168–173.

Huchra, J., Gorenstein, M., Kent, S., Shapiro, I., Smith, G., Horine, E. and Perley, R. (1985). 2237 + 0305: A new and unusual gravitational lens. *AJ*, 90:691–696.

Hughes, J.P. and Birkinshaw, M. (1998). A Measurement of the Hubble Constant from the X-Ray Properties and the Sunyaev-Zeldovich Effect of CL 0016+16. *ApJ*, 501:1–14.

Hui, L. and Greene, P.B. (2006). Correlated fluctuations in luminosity distance and the importance of peculiar motion in supernova surveys. *Phys. Rev. D*, 73(12):123526.

Humphreys, E.M.L., Reid, M.J., Moran, J.M., Greenhill, L.J. and Argon, A.L. (2013). Toward a New Geometric Distance to the Active Galaxy NGC 4258. III. Final Results and the Hubble Constant. *ApJ*, 775:13.

Impey, C.D., Falco, E.E., Kochanek, C.S., Lehár, J., McLeod, B.A., Rix, H.W., Peng, C.Y. and Keeton, C.R. (1998). An Infrared Einstein Ring in the Gravitational Lens PG 1115+080. *ApJ*, 509:551–560.

Jang, I.S., Hatt, D., Beaton, R.L., Lee, M.G., Freedman, W.L. et al (2017). The Carnegie-Chicago Hubble Program. III. The Distance to NGC 1365 via the Tip of the Red Giant Branch. *ArXiv e-prints:1703.10616*.

Jee, I., Komatsu, E., Suyu, S.H. and Huterer, D. (2016). Time-delay cosmography: increased leverage with angular diameter distances. *J. Cosmology Astropart. Phys.*, 4:031.

Jeffreys, H. (1998). *The Theory of Probability*. Oxford Classic Texts in the Physical Sciences. OUP Oxford.

Joseph, R., Courbin, F., Metcalf, R.B., Giocoli, C., Hartley, P. et al (2014). A PCA-based automated finder for galaxy-scale strong lenses. *A&A*, 566:A63.

Kaiser, N., Wilson, G. and Luppino, G.A. (2000). Large-Scale Cosmic Shear Measurements. *ArXiv Astrophysics e-prints*.

Kapteyn, J.C. (1922). First Attempt at a Theory of the Arrangement and Motion of the Sidereal System. *ApJ*, 55:302.

Karwal, T. and Kamionkowski, M. (2016). Dark energy at early times, the Hubble parameter, and the string axiverse. *Phys. Rev. D*, 94(10):103523.

Keeton, C.R. and Zabludoff, A.I. (2004). The Importance of Lens Galaxy Environments. *ApJ*, 612:660–678.

Keisler, R., Reichardt, C.L., Aird, K.A., Benson, B.A., Bleem, L.E. et al (2011). A Measurement of the Damping Tail of the Cosmic Microwave Background Power Spectrum with the South Pole Telescope. *ApJ*, 743:28.

Kellermann, K.I. (2013). The discovery of quasars. *Bulletin of the Astronomical Society of India*, 41:1.

Kelly, B.C., Becker, A.C., Sobolewska, M., Siemiginowska, A. and Uttley, P. (2014). Flexible and Scalable Methods for Quantifying Stochastic Variability in the Era of Massive Time-domain Astronomical Data Sets. *ApJ*, 788:33.

Kelly, P.L., Rodney, S.A., Treu, T., Foley, R.J., Brammer, G. et al (2015). Multiple images of a highly magnified supernova formed by an early-type cluster galaxy lens. *Science*, 347:1123–1126.

Kelly, P.L., Rodney, S.A., Treu, T., Strolger, L.G., Foley, R.J. et al (2016). Deja Vu All Over Again: The Reappearance of Supernova Refsdal. *ApJ*, 819:L8.

Kirshner, R.P. (2004). Hubble's diagram and cosmic expansion. *Proceedings of the National Academy of Sciences of the United States of America*, 101(1):8–13.

Klypin, A., Kravtsov, A.V., Valenzuela, O. and Prada, F. (1999). Where Are the Missing Galactic Satellites? *ApJ*, 522:82–92.

Kochanek, C.S. (2004). Quantitative Interpretation of Quasar Microlensing Light Curves. *ApJ*, 605:58–77.

Kochanek, C.S., Dai, X., Morgan, C., Morgan, N. and Poindexter, G., S.C. (2007). Turning AGN Microlensing from a Curiosity into a Tool. In Babu, G.J. and Feigelson, E.D., editors, *Statistical Challenges in Modern Astronomy IV*, volume 371 of *Astronomical Society of the Pacific Conference Series*, page 43.

Kochanek, C.S., Falco, E.E., Impey, C., Lehar, J., McLeod, B. and Rix, H.W. (1998). CfA-Arizona Space Telescope LEns Survey. [Online; accessed 01-July-2017].

Kochanek, C.S., Morgan, N.D., Falco, E.E., McLeod, B.A., Winn, J.N., Dembicky, J. and Ketzeback, B. (2006). The Time Delays of Gravitational Lens HE 0435-1223: An Early-Type Galaxy with a Rising Rotation Curve. *ApJ*, 640:47–61.

Kochanek, C.S., Shappee, B.J., Stanek, K.Z., Holoien, T.W.S., Thompson, T.A. et al (2017). The All-Sky Automated Survey for Supernovae (ASAS-SN) Light Curve Server v1.0. *ArXiv e-prints:1706.07060*.

Kogut, A., Lineweaver, C., Smoot, G.F., Bennett, C.L., Banday, A. et al (1993). Dipole Anisotropy in the COBE Differential Microwave Radiometers First-Year Sky Maps. *ApJ*, 419:1.

Köhlinger, F., Viola, M., Joachimi, B., Hoekstra, H., van Uitert, E. et al (2017). KiDS-450: The tomographic weak lensing power spectrum and constraints on cosmological parameters. *ArXiv e-prints:1706.02892*.

Koopmans, L.V.E. and Treu, T. (2003). The Structure and Dynamics of Luminous and Dark Matter in the Early-Type Lens Galaxy of 0047-281 at z = 0.485. *ApJ*, 583:606–615.

Kosowsky, A. (2003). The Atacama Cosmology Telescope. *New A Rev.*, 47:939–943.

Kowal, C.T. (1968). Absolute magnitudes of supernovae. *AJ*, 73:1021–1024.

Kravtsov, A. (2010). The Dark Matter Annihilation Signal from Dwarf Galaxies and Subhalos. *Advances in Astronomy*, 2010:281913.

Kristian, J., Groth, E.J., Shaya, E.J., Schneider, D.P., Holtzman, J.A. et al (1993). Imaging of the gravitational lens system PG 1115+080 with the Hubble Space Telescope. *AJ*, 106:1330–1336.

Kundic, T., Cohen, J.G., Blandford, R.D. and Lubin, L.M. (1997). Keck Spectroscopy of the Gravitational Lens System PG 1115+080: Redshifts of the Lensing Galaxies. *AJ*, 114:507–510.

Kundić, T., Turner, E.L., Colley, W.N., Gott, III, J.R., Rhoads, J.E. et al (1997). A Robust Determination of the Time Delay in 0957+561A, B and a Measurement of the Global Value of Hubble's Constant. *ApJ*, 482:75–82.

Kuo, C.Y., Braatz, J.A., Reid, M.J., Lo, K.Y., Condon, J.J., Impellizzeri, C.M.V. and Henkel, C. (2013). The Megamaser Cosmology Project. V. An Angular-diameter Distance to NGC 6264 at 140 Mpc. *ApJ*, 767:155.

Lanusse, F., Ma, Q., Li, N., Collett, T.E., Li, C.L., Ravanbakhsh, S., Mandelbaum, R. and Poczos, B. (2017). CMU DeepLens: Deep Learning For Automatic Image-based Galaxy-Galaxy Strong Lens Finding. *ArXiv e-prints:1703.02642*.

Larson, D., Weiland, J.L., Hinshaw, G. and Bennett, C.L. (2015). Comparing Planck and WMAP: Maps, Spectra, and Parameters. *ApJ*, 801:9.

Laureijs, R., Amiaux, J., Arduini, S., Augueres, J.., Brinchmann, J. et al (2011). Euclid Definition Study Report. *ArXiv e-prints:1110.3193*.

Leavitt, H.S. (1908). 1777 variables in the Magellanic Clouds. *Annals of Harvard College Observatory*, 60:87–108.3.

Lee, M.G., Freedman, W.L. and Madore, B.F. (1993). The Tip of the Red Giant Branch as a Distance Indicator for Resolved Galaxies. *ApJ*, 417:553.

Lehar, J., Hewitt, J.N., Burke, B.F. and Roberts, D.H. (1992). The radio time delay in the double quasar 0957 + 561. *ApJ*, 384:453–466.

Leibundgut, B. (2001). Cosmological Implications from Observations of Type Ia Supernovae. *ARA&A*, 39:67–98.

Lemaître, G. (1927). Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. *Annales de la Société Scientifique de Bruxelles*, 47:49–59.

Lemaître, G. (1931a). Expansion of the universe, The expanding universe. *MNRAS*, 91:490–501.

Lemaître, G. (1931b). The Beginning of the World from the Point of View of Quantum Theory. *Nature*, 127:706.

Lewis, A. and Bridle, S. (2002). Cosmological parameters from CMB and other data: A Monte Carlo approach. *Phys. Rev. D*, 66(10):103511.

Li, W., Filippenko, A.V., Chornock, R., Berger, E., Berlind, P. et al (2003). SN 2002cx: The Most Peculiar Known Type Ia Supernova. *PASP*, 115:453–473.

Liao, K., Treu, T., Marshall, P., Fassnacht, C.D., Rumbaugh, N. et al (2015). Strong Lens Time Delay Challenge. II. Results of TDC1. *ApJ*, 800:11.

Lidman, C., Courbin, F., Kneib, J.P., Golse, G., Castander, F. and Soucail, G. (2000). Exploring the gravitationally lensed system HE 1104-1805: VLT spectroscopy of the lens at z=0.729. *A&A*, 364:L62–L65.

Liesenborgs, J. and De Rijcke, S. (2012). Lensing degeneracies and mass substructure. *MNRAS*, 425:1772–1780.

Lin, H., Buckley-Geer, E., Agnello, A., Ostrovski, F., McMahon, R.G. et al (2017). Discovery of the Lensed Quasar System DES J0408-5354. *ArXiv e-prints:1702.00072*.

Linder, E.V. (2011). Lensing time delays and cosmological complementarity. *Phys. Rev. D*, 84(12):123529.

Linder, E.V. (2015). Tailoring strong lensing cosmographic observations. *Phys. Rev. D*, 91(8):083511.

Liske, J., Baldry, I.K., Driver, S.P., Tuffs, R.J., Alpaslan, M. et al (2015). Galaxy And Mass Assembly (GAMA): end of survey report and data release 2. *MNRAS*, 452:2087–2126.

Livio, M. and Riess, A.G. (2013). Measuring the Hubble constant. *Physics Today*, 41(66).

Lovell, J.E.J., Jauncey, D.L., Reynolds, J.E., Wieringa, M.H., King, E.A., Tzioumis, A.K., McCulloch, P.M. and Edwards, P.G. (1998). The Time Delay in the Gravitational Lens PKS 1830-211. *ApJ*, 508:L51–L54.

LSST Science Collaboration, Abell, P.A., Allison, J., Anderson, S.F., Andrew, J.R. et al (2009). LSST Science Book, Version 2.0. *ArXiv e-prints:0912.0201*.

Lynds, R. and Petrosian, V. (1986). Giant Luminous Arcs in Galaxy Clusters. In *Bulletin of the American Astronomical Society*, volume 18 of *BAAS*, page 1014.

Maccoun, R. and Perlmutter, S. (2015). Blind analysis: Hide results to seek the truth. *Nature*, 526:187–189.

Macri, L.M. and Riess, A.G. (2009). The SH0ES Project: Observations of Cepheids in NGC 4258 and Type Ia SN Hosts. In Guzik, J.A. and Bradley, P.A., editors, *American Institute of Physics Conference Series*, volume 1170 of *American Institute of Physics Conference Series*, pages 23–25.

Madore, B.F. and Freedman, W.L. (1991). The Cepheid distance scale. *PASP*, 103:933–957.

Magain, P., Courbin, F. and Sohy, S. (1998). Deconvolution with Correct Sampling. *ApJ*, 494:472–477.

Magain, P., Surdej, J., Swings, J.P., Borgeest, U. and Kayser, R. (1988). Discovery of a quadruply lensed quasar - The ’clover leaf’ H1413 + 117. *Nature*, 334:325–327.

Mao, S. and Paczynski, B. (1991). Gravitational microlensing by double stars and planetary systems. *ApJ*, 374:L37–L40.

Mao, S. and Schneider, P. (1998). Evidence for substructure in lens galaxies? *MNRAS*, 295:587.

Marshall, P., Rajguru, N. and Slosar, A. (2006). Bayesian evidence as a tool for comparing datasets. *Phys. Rev. D*, 73(6):067302.

Marshall, P.J., Treu, T., Melbourne, J., Gavazzi, R., Bundy, K. et al (2007). Superresolving Distant Galaxies with Gravitational Telescopes: Keck Laser Guide Star Adaptive Optics and Hubble Space Telescope Imaging of the Lens System SDSS J0737+3216. *ApJ*, 671:1196–1211.

Marshall, P.J., Verma, A., More, A., Davis, C.P., More, S. et al (2016). SPACE WARPS - I. Crowdsourcing the discovery of gravitational lenses. *MNRAS*, 455:1171–1190.

McCully, C., Keeton, C.R., Wong, K.C. and Zabludoff, A.I. (2014). A new hybrid framework to efficiently model lines of sight to gravitational lenses. *MNRAS*, 443:3631–3642.

McCully, C., Keeton, C.R., Wong, K.C. and Zabludoff, A.I. (2017). Quantifying Environmental and Line-of-sight Effects in Models of Strong Gravitational Lens Systems. *ApJ*, 836:141.

Mediavilla, E., Jiménez-Vicente, J., Muñoz, J.A., Vives-Arias, H. and Calderón-Infante, J. (2017). Limits on the Mass and Abundance of Primordial Black Holes from Quasar Gravitational Microlensing. *ApJ*, 836:L18.

Mediavilla, E., Muñoz, J.A., Falco, E., Motta, V., Guerras, E. et al (2009). Microlensing-based Estimate of the Mass Fraction in Compact Objects in Lens Galaxies. *ApJ*, 706:1451–1462.

Meng, X.L., Treu, T., Agnello, A., Auger, M.W., Liao, K. and Marshall, P.J. (2015). Precision cosmology with time delay lenses: high resolution imaging requirements. *J. Cosmology Astropart. Phys.*, 9:059.

Meylan, G., Jetzer, P., North, P., Schneider, P., Kochanek, C.S. and Wambsganss, J., editors (2006). *Gravitational Lensing: Strong, Weak and Micro.*

Michell, J. (1784). On the Means of Discovering the Distance, Magnitude, of the Fixed Stars, in Consequence of the Diminution of the Velocity of Their Light, in Case Such a Diminution Should be Found to Take Place in any of Them, and Such Other Data Should be Procured from Observations, as Would be Farther Necessary for That Purpose. By the Rev. John Michell, B. D. F. R. S. In a Letter to Henry Cavendish, Esq. F. R. S. and A. S. *Philosophical Transactions of the Royal Society of London Series I*, 74:35–57.

Moehler, S., Freudling, W., Møller, P., Patat, F., Rupprecht, G. and O'Brien, K. (2010). Correction of Field Rotator-Induced Flat-Field Systematics - A Case Study Using Archived VLT-FORS Data. *PASP*, 122:93.

Molinari, N., Durand, J. and Sabatier, R. (2004). Bounded optimal knots for regression splines. *Computational Statistics & Data Analysis*, 45(2):159–178.

Momcheva, I., Williams, K., Keeton, C. and Zabludoff, A. (2006). A Spectroscopic Study of the Environments of Gravitational Lens Galaxies. *ApJ*, 641:169–189.

Moore, B., Ghigna, S., Governato, F., Lake, G., Quinn, T., Stadel, J. and Tozzi, P. (1999). Dark Matter Substructure within Galactic Halos. *ApJ*, 524:L19–L22.

More, A., Oguri, M., Kayo, I., Zinn, J., Strauss, M.A. et al (2016a). The SDSS-III BOSS quasar lens survey: discovery of 13 gravitationally lensed quasars. *MNRAS*, 456:1595–1606.

More, A., Verma, A., Marshall, P.J., More, S., Baeten, E. et al (2016b). SPACE WARPS- II. New gravitational lens candidates from the CFHTLS discovered through citizen science. *MNRAS*, 455:1191–1210.

Morgan, C.W., Hainline, L.J., Chen, B., Tewes, M., Kochanek, C.S. et al (2012). Further Evidence that Quasar X-Ray Emitting Regions are Compact: X-Ray and Optical Microlensing in the Lensed Quasar Q J0158-4325. *ApJ*, 756:52.

Morgan, C.W., Kochanek, C.S., Morgan, N.D. and Falco, E.E. (2010). The Quasar Accretion Disk Size-Black Hole Mass Relation. *ApJ*, 712:1129–1136.

Morgan, N.D., Caldwell, J.A.R., Schechter, P.L., Dressler, A., Egami, E. and Rix, H.W. (2004). WFI J2026-4536 and WFI J2033-4723: Two New Quadruple Gravitational Lenses. *AJ*, 127:2617–2630.

Morgan, N.D., Kochanek, C.S., Falco, E.E. and Dai, X. (2006). Time-Delay Measurement for the Quadruple Lens RX J1131-1231. *ArXiv Astrophysics e-prints*.

Morgan, N.D., Kochanek, C.S., Pevunova, O. and Schechter, P.L. (2005). The Lens Redshift and Galaxy Environment for HE 0435-1223. *AJ*, 129:2531–2541.

Mushotzky, R.F., Edelson, R., Baumgartner, W. and Gandhi, P. (2011). Kepler Observations of Rapid Optical Variability in Active Galactic Nuclei. *ApJ*, 743:L12.

Myers, S.T., Fassnacht, C.D., Djorgovski, S.G., Blandford, R.D., Matthews, K. et al (1995). 1608+656: A Quadruple-Lens System Found in the CLASS Gravitational Lens Survey. *ApJ*, 447:L5.

Nakajima, R., Bernstein, G.M., Fadely, R., Keeton, C.R. and Schrabback, T. (2009). Improved Constraints on the Gravitational Lens Q0957+561. I. Weak Lensing. *ApJ*, 697:1793–1804.

Ofek, E.O. and Maoz, D. (2003). Time-Delay Measurement of the Lensed Quasar HE 1104-1805. *ApJ*, 594:101–106.

Ofek, E.O., Maoz, D., Rix, H.W., Kochanek, C.S. and Falco, E.E. (2006). Spectroscopic Redshifts for Seven Lens Galaxies. *ApJ*, 641:70–77.

Oguri, M. (2007). Gravitational Lens Time Delays: A Statistical Assessment of Lens Model Dependences and Implications for the Global Hubble Constant. *ApJ*, 660:1–15.

Oguri, M. (2015). Predicted properties of multiple images of the strongly lensed supernova SN Refsdal. *MNRAS*, 449:L86–L89.

Oguri, M. and Marshall, P.J. (2010). Gravitationally lensed quasars and supernovae in future wide-field optical imaging surveys. *MNRAS*, 405:2579–2593.

Oscoz, A., Mediavilla, E., Goicoechea, L.J., Serra-Ricart, M. and Buitrago, J. (1997). Time Delay of QSO 0957+561 and Cosmological Implications. *ApJ*, 479:L89–L92.

Paczynski, B. (1986). Gravitational microlensing by the galactic halo. *ApJ*, 304:1–5.

Paczynski, B. (1991). Gravitational microlensing of the Galactic bulge stars. *ApJ*, 371:L63–L67.

Paczynski, B. (1997). Detached eclipsing binaries as primary distance and age indicators. In Livio, M., Donahue, M. and Panagia, N., editors, *The Extragalactic Distance Scale*, pages 273–280.

Paraficz, D., Hjorth, J. and Elíasdóttir, Á. (2009). Results of optical monitoring of 5 SDSS double QSOs with the Nordic Optical Telescope. *A&A*, 499:395–408.

Paraficz, D., Rybak, M., McKean, J.P., Vegetti, S., Sluse, D. et al (2017). ALMA view of RX J1131-1231: Sub-kpc CO (2-1) mapping of a molecular disk in a lensed star-forming quasar host galaxy. *ArXiv e-prints:1705.09931*.

Patil, A., Huard, D. and Fonnesbeck, C.J. (2010). PyMC: Bayesian Stochastic Modelling in Python. *Journal of statistical software*, 35(4):1–81.

Patnaik, A.R., Browne, I.W.A., Walsh, D., Chaffee, F.H. and Foltz, C.B. (1992). B1422+231 - A new gravitationally lensed system at Z = 3.62. *MNRAS*, 259:1P–4P.

Peebles, P.J.E. (1968). Recombination of the Primeval Plasma. *ApJ*, 153:1.

Peebles, P.J.E. and Yu, J.T. (1970). Primeval Adiabatic Perturbation in an Expanding Universe. *ApJ*, 162:815.

Pelt, J., Kayser, R., Refsdal, S. and Schramm, T. (1996). The light curve and the time delay of QSO 0957+561. *A&A*, 305:97.

Pelt, J., Refsdal, S. and Stabell, R. (2002). Bias and consistency in time delay estimation methods: Case of the double quasar HE 1104-1805. *A&A*, 389:L57–L60.

Pelt, J., Schild, R., Refsdal, S. and Stabell, R. (1998). Microlensing on different timescales in the lightcurves of QSO 0957+561 A,B. *A&A*, 336:829–839.

Penzias, A.A. and Wilson, R.W. (1965). A Measurement of Excess Antenna Temperature at 4080 Mc/s. *ApJ*, 142:419–421.

Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R.A., Nugent, P. et al (1999). Measurements of Ω and Λ from 42 High-Redshift Supernovae. *ApJ*, 517:565–586.

Petrillo, C.E., Tortora, C., Chatterjee, S., Vernardos, G., Koopmans, L.V.E. et al (2017). Finding Strong Gravitational Lenses in the Kilo Degree Survey with Convolutional Neural Networks. *ArXiv e-prints:1702.07675*.

Phillips, M.M. (1993). The absolute magnitudes of Type IA supernovae. *ApJ*, 413:L105–L108.

Planck Collaboration (2005). *The Scientific Programme of Planck (Bluebook).*

Planck Collaboration, Adam, R., Ade, P.A.R., Aghanim, N., Akrami, Y. et al (2016a). Planck 2015 results. I. Overview of products and scientific results. *A&A*, 594:A1.

Planck Collaboration, Adam, R., Ade, P.A.R., Aghanim, N., Arnaud, M. et al (2016b). Planck intermediate results. XXX. The angular power spectrum of polarized dust emission at intermediate and high Galactic latitudes. *A&A*, 586:A133.

Planck Collaboration, Ade, P.A.R., Aghanim, N., Arnaud, M., Ashdown, M. et al (2016c). Planck 2015 results. XIII. Cosmological parameters. *A&A*, 594:A13.

Planck Collaboration, Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M. et al (2014). Planck 2013 results. XXVII. Doppler boosting of the CMB: Eppur si muove. *A&A*, 571:A27.

Planck Collaboration, Aghanim, N., Ashdown, M., Aumont, J., Baccigalupi, C. et al (2016d). Planck intermediate results. XLVI. Reduction of large-scale systematic effects in HFI polarization maps and estimation of the reionization optical depth. *A&A*, 596:A107.

Poindexter, S. and Kochanek, C.S. (2010a). Microlensing Evidence that a Type 1 Quasar is Viewed Face-On. *ApJ*, 712:668–673.

Poindexter, S. and Kochanek, C.S. (2010b). The Transverse Peculiar Velocity of the Q2237+0305 Lens Galaxy and the Mean Mass of Its Stars. *ApJ*, 712:658–667.

Poindexter, S., Morgan, N. and Kochanek, C.S. (2008). The Spatial Structure of an Accretion Disk. *ApJ*, 673:34–38.

Poindexter, S., Morgan, N., Kochanek, C.S. and Falco, E.E. (2007). Mid-IR Observations and a Revised Time Delay for the Gravitational Lens System Quasar HE 1104-1805. *ApJ*, 660:146–151.

Press, W.H., Rybicki, G.B. and Hewitt, J.N. (1992a). The time delay of gravitational lens 0957 + 561. I - Methodology and analysis of optical photometric data. II - Analysis of radio data and combined optical-radio analysis. *ApJ*, 385:404–420.

Press, W.H., Rybicki, G.B. and Hewitt, J.N. (1992b). The Time Delay of Gravitational Lens 0957+561. II. Analysis of Radio Data and Combined Optical-Radio Analysis. *ApJ*, 385:416.

Rasmussen, C.E. and Williams, C.K.I. (2006). *Gaussian processes for machine learning.*, volume 14.

Rathna Kumar, S. (2017). Crash testing difference-smoothing algorithm on a large sample of simulated light curves from TDC1. *ArXiv e-prints:1704.00706.*

Rathna Kumar, S., Tewes, M., Stalin, C.S., Courbin, F., Asfandiyarov, I. et al (2013). COSMOGRAIL: the COSmological MOnitoring of GRAvItational Lenses. XIV. Time delay of the doubly lensed quasar SDSS J1001+5027. *A&A*, 557:A44.

Refsdal, S. (1964a). On the possibility of determining Hubble's parameter and the masses of galaxies from the gravitational lens effect. *MNRAS*, 128:307.

Refsdal, S. (1964b). The gravitational lens effect. *MNRAS*, 128:295.

Refsdal, S., Stabell, R., Pelt, J. and Schild, R. (2000). Constraints on source and lens parameters from microlensing variability in QSO 0957+561 A,B. *A&A*, 360:10–14.

Reid, M.J., Braatz, J.A., Condon, J.J., Lo, K.Y., Kuo, C.Y., Impellizzeri, C.M.V. and Henkel, C. (2013). The Megamaser Cosmology Project. IV. A Direct Measurement of the Hubble Constant from UGC 3789. *ApJ*, 767:154.

Reimers, D., Koehler, T. and Wisotzki, L. (1996). The Hamburg/ESO survey for bright QSOs. II. Follow-up spectroscopy of 160 quasars and Seyferts. *A&AS*, 115:235.

Remy, M., Claeskens, J.F., Surdej, J., Hjorth, J., Refsdal, S., Wucknitz, O., Sørensen, A.N. and Grundahl, F. (1998). Detection of the lensing galaxy for the double QSO HE 1104-1805. *New A*, 3:379–390.

Rhee, G. (1991). An estimate of the Hubble constant from the gravitational lensing of quasar Q0957 + 561. *Nature*, 350:211.

Riess, A.G., Filippenko, A.V., Challis, P., Clocchiatti, A., Diercks, A. et al (1998). Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *AJ*, 116:1009–1038.

Riess, A.G., Macri, L., Casertano, S., Lampeitl, H., Ferguson, H.C. et al (2011). A 3% Solution: Determination of the Hubble Constant with the Hubble Space Telescope and Wide Field Camera 3. *ApJ*, 730:119.

Riess, A.G., Macri, L.M., Hoffmann, S.L., Scolnic, D., Casertano, S. et al (2016). A 2.4% Determination of the Local Value of the Hubble Constant. *ApJ*, 826:56.

Rigault, M., Aldering, G., Kowalski, M., Copin, Y., Antilogus, P. et al (2015). Confirmation of a Star Formation Bias in Type Ia Supernova Distances and its Effect on the Measurement of the Hubble Constant. *ApJ*, 802:20.

Rubin, V.C. and Ford, Jr., W.K. (1970). Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. *ApJ*, 159:379.

Rusu, C.E., Fassnacht, C.D., Sluse, D., Hilbert, S., Wong, K.C. et al (2017). H0LiCOW - III. Quantifying the effect of mass along the line of sight to the gravitational lens HE 0435-1223 through weighted galaxy counts. *MNRAS*, 467:4220–4242.

Saha, P., Courbin, F., Sluse, D., Dye, S. and Meylan, G. (2006). COSMOGRAIL: the COSmological MOnitoring of GRAvItational Lenses. IV. Models of prospective time-delay lenses. *A&A*, 450:461–469.

Sandage, A. (1958). Current Problems in the Extragalactic Distance Scale. *ApJ*, 127:513.

Sandage, A. (1961). The Ability of the 200-INCH Telescope to Discriminate Between Selected World Models. *ApJ*, 133:355.

Sandage, A. and Tammann, G.A. (1975). Steps toward the Hubble constant. V - The Hubble constant from nearby galaxies and the regularity of the local velocity field. *ApJ*, 196:313–328.

Schaefer, C., Geiger, M., Kuntzer, T. and Kneib, J. (2017). Deep Convolutional Neural Networks as strong gravitational lens detectors. *ArXiv e-prints:1705.07132*.

Schechter, P.L., Bailyn, C.D., Barr, R., Barvainis, R., Becker, C.M. et al (1997). The Quadruple Gravitational Lens PG 1115+080: Time Delays and Models. *ApJ*, 475:L85–L88.

Schechter, P.L. and Wambsganss, J. (2002). Quasar Microlensing at High Magnification and the Role of Dark Matter: Enhanced Fluctuations and Suppressed Saddle Points. *ApJ*, 580:685–695.

Schild, R. and Thomson, D.J. (1995). Twin QSO Q0957+561 Time Delay Dataset. *AJ*, 109:1970.

Schild, R.E. (1990). The time delay in the twin QSO Q0957 + 561. *AJ*, 100:1771–1776.

Schild, R.E. and Smith, R.C. (1991). Microlensing in the Q0957 + 561 gravitational mirage. *AJ*, 101:813–817.

Schmidt, B.P., Suntzeff, N.B., Phillips, M.M., Schommer, R.A., Clocchiatti, A. et al (1998). The High-Z Supernova Search: Measuring Cosmic Deceleration and Global Curvature of the Universe Using Type IA Supernovae. *ApJ*, 507:46–63.

Schmidt, M. (1963). 3C 273 : A Star-Like Object with Large Red-Shift. *Nature*, 197:1040.

Schneider, P. (1985). A new formulation of gravitational lens theory, time-delay, and Fermat's principle. *A&A*, 143:413–420.

Schneider, P. (2006). Gravitational Lensing: An Overview. In *KITP: Blackboard Lunch Series*.

Schneider, P. (2014). Generalized multi-plane gravitational lensing: time delays, recursive lens equation, and the mass-sheet transformation. *ArXiv e-prints:1409.0015*.

Schneider, P., Ehlers, J. and Falco, E.E. (1992). *Gravitational Lenses*.

Schneider, P. and Sluse, D. (2013). Mass-sheet degeneracy, power-law models and external convergence: Impact on the determination of the Hubble constant from gravitational lensing. *A&A*, 559:A37.

Schneider, P. and Sluse, D. (2014). Source-position transformation: an approximate invariance in strong gravitational lensing. *A&A*, 564:A103.

Shanks, T., Metcalfe, N., Chehade, B., Findlay, J.R., Irwin, M.J. et al (2015). The VLT Survey Telescope ATLAS. *MNRAS*, 451:4238–4252.

Shields, G.A. (1999). A Brief History of Active Galactic Nuclei. *P.A.S.P*, 111:661–678.

Shimanovskaya, E.V., Oknyanskii, V.L. and Artamonov, B.P. (2015). Determining the time delays in the gravitational lens PG 1115+080. *Astronomy Reports*, 59:12–24.

Silk, J. and White, S.D.M. (1978). The determination of Q_0 using X-ray and microwave observations of galaxy clusters. *ApJ*, 226:L103–L106.

Simpson, F., Jimenez, R., Pena-Garay, C. and Verde, L. (2017). Strong Evidence for the Normal Neutrino Hierarchy. *ArXiv e-prints:1703.03425*.

Sluse, D., Chantry, V., Magain, P., Courbin, F. and Meylan, G. (2012a). COSMOGRAIL: the COSmological MOnitoring of GRAvItational Lenses. X. Modeling based on high-precision astrometry of a sample of 25 lensed quasars: consequences for ellipticity, shear, and astrometric anomalies. *A&A*, 538:A99.

Sluse, D., Hutsemékers, D., Courbin, F., Meylan, G. and Wambsganss, J. (2012b). Microlensing of the broad line region in 17 lensed quasars. *A&A*, 544:A62.

Sluse, D., Hutsemékers, D., Courbin, F., Meylan, G. and Wambsganss, J. (2012c). Microlensing of the broad line region in 17 lensed quasars. *A&A*, 544:A62.

Sluse, D., Schmidt, R., Courbin, F., Hutsemékers, D., Meylan, G. et al (2011). Zooming into the broad line region of the gravitationally lensed quasar QSO 2237 + 0305, the Einstein Cross. III. Determination of the size and structure of the C iv and C iii] emitting regions using microlensing. *A&A*, 528:A100.

Sluse, D., Sonnenfeld, A., Rumbaugh, N., Rusu, C.E., Fassnacht, C.D. et al (2016). H0LiCOW II. Spectroscopic survey and galaxy-group identification of the strong gravitational lens system HE0435-1223. *ArXiv e-prints:1607.00382*.

Sluse, D., Surdej, J., Claeskens, J.F., Hutsemékers, D., Jean, C. et al (2003). A quadruply imaged quasar with an optical Einstein ring candidate: 1RXS J113155.4-123155. *A&A*, 406:L43–L46.

Smette, A., Robertson, J.G., Shaver, P.A., Reimers, D., Wisotzki, L. and Koehler, T. (1995). The gravitational lens candidate HE 1104-1805 and the size of absorption systems. *A&AS*, 113:199.

Smoot, G.F., Bennett, C.L., Kogut, A., Wright, E.L., Aymon, J. et al (1992a). Structure in the COBE differential microwave radiometer first-year maps. *ApJ*, 396:L1–L5.

Smoot, G.F., Bennett, C.L., Kogut, A., Wright, E.L., Aymon, J. et al (1992b). Structure in the COBE differential microwave radiometer first-year maps. *ApJ*, 396:L1–L5.

Snellen, I.A.G., de Bruyn, A.G., Schilizzi, R.T., Miley, G.K. and Myers, S.T. (1995). Radio Observations of the Quadruple Lens 1608+656. *ApJ*, 447:L9.

Sokol, J. (2017). Hubble trouble. *Science*, 355(6329):1010–1014.

Soldner, J. (1801). *Berliner astronomisches Jahrbuch für...* F. Dümmler.

Sonnenfeld, A., Chan, J.H.H., Shu, Y., More, A., Oguri, M. et al (2017). Survey of Gravitationally-lensed Objects in HSC Imaging (SuGOHI). I. Automatic search for galaxy-scale strong lenses. *ArXiv e-prints:1704.01585.*

Sorce, J.G., Tully, R.B. and Courtois, H.M. (2012). The Mid-infrared Tully-Fisher Relation: Calibration of the Type Ia Supernova Scale and H $_0$. *ApJ*, 758:L12.

Soucail, G., Fort, B., Mellier, Y. and Picat, J.P. (1987). A blue ring-like structure, in the center of the {A} 370 cluster of galaxies. *a & a*, 172:L14—-L16.

Spergel, D., Gehrels, N., Baltay, C., Bennett, D., Breckinridge, J. et al (2015a). Wide-Field InfrarRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report. *ArXiv e-prints:1503.03757.*

Spergel, D.N., Flauger, R. and Hložek, R. (2015b). Planck data reconsidered. *Phys. Rev. D*, 91(2):023518.

Springel, V., White, S.D.M., Jenkins, A., Frenk, C.S., Yoshida, N. et al (2005). Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 435:629–636.

Stromberg, G. (1925). No. 292. Analysis of radial velocities of globular clusters and non-galactic nebulae. *Contributions from the Mount Wilson Observatory / Carnegie Institution of Washington*, 292:1–10.

Sunyaev, R.A. and Zeldovich, Y.B. (1972). The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies. *Comments on Astrophysics and Space Physics*, 4:173.

Suyu, S.H., Auger, M.W., Hilbert, S., Marshall, P.J., Tewes, M. et al (2013). Two Accurate Time-delay Distances from Strong Lensing: Implications for Cosmology. *ApJ*, 766:70.

Suyu, S.H., Bonvin, V., Courbin, F., Fassnacht, C.D., Rusu, C.E. et al (2016). H0LiCOW I. *H*_0 Lenses in COSMOGRAIL's Wellspring: Program Overview. *ArXiv e-prints:1607.00017.*

Suyu, S.H. and Halkola, A. (2010). The halos of satellite galaxies: the companion of the massive elliptical lens SL2S J08544-0121. *A&A*, 524:A94.

Suyu, S.H., Hensel, S.W., McKean, J.P., Fassnacht, C.D., Treu, T. et al (2012). Disentangling Baryons and Dark Matter in the Spiral Gravitational Lens B1933+503. *ApJ*, 750:10.

Suyu, S.H., Marshall, P.J., Auger, M.W., Hilbert, S., Blandford, R.D., Koopmans, L.V.E., Fassnacht, C.D. and Treu, T. (2010). Dissecting the Gravitational lens B1608+656. II. Precision Measurements of the Hubble Constant, Spatial Curvature, and the Dark Energy Equation of State. *ApJ*, 711:201–221.

Suyu, S.H., Marshall, P.J., Hobson, M.P. and Blandford, R.D. (2006). A Bayesian analysis of regularized source inversions in gravitational lensing. *MNRAS*, 371:983–998.

Suyu, S.H., Treu, T., Hilbert, S., Sonnenfeld, A., Auger, M.W. et al (2014). Cosmology from Gravitational Lens Time Delays and Planck Data. *ApJ*, 788:L35.

Szabo, T. (2014). DIFFERENT PERSPECTIVES. [Online; accessed 01-July-2017].

Tagore, A.S., Barnes, D.J., Jackson, N., Kay, S.T., Schaller, M., Schaye, J. and Theuns, T. (2017). Reducing biases on H_0 measurements using strong lensing and galaxy dynamics: results from the EAGLE simulation. *ArXiv e-prints:1706.07733*.

Tak, H., Mandel, K., van Dyk, D.A., Kashyap, V.L., Meng, X.L. and Siemiginowska, A. (2016). Bayesian Estimates of Astronomical Time Delays between Gravitationally Lensed Stochastic Light Curves. *ArXiv e-prints:1602.01462*.

Tammann, G.A. and Leibundgut, B. (1990). Supernova studies. IV - The global value of H0 from supernovae IA and the peculiar motion of field galaxies. *A&A*, 236:9–14.

Tewes, M., Courbin, F. and Meylan, G. (2013a). COSMOGRAIL: the COSmological MOnitoring of GRAvItational Lenses. XI. Techniques for time delay measurement in presence of microlensing. *A&A*, 553:A120.

Tewes, M., Courbin, F., Meylan, G., Kochanek, C.S., Eulaers, E. et al (2013b). COSMOGRAIL: the COSmological MOnitoring of GRAvItational Lenses. XIII. Time delays and 9-yr optical monitoring of the lensed quasar RX J1131-1231. *A&A*, 556:A22.

Tewes, M., Meylan, G. and Courbin, F. (2013c). Cosmology with Gravitational Lensing: Measuring Quasar Time Delays and Cosmic Shear. *EPFL PhD Thesis*, page 706.

Tie, S.S. and Kochanek, C.S. (2017). Microlensing Makes Lensed Quasar Time Delays Significantly Time Variable. *ArXiv e-prints:1707.01908*.

Timmer, J. and Koenig, M. (1995). On generating power law noise. *A&A*, 300:707.

Tonry, J.L. (1998). Redshifts of the gravitational lenses B1422+231 and PG 1115+080. *AJ*, 115:1.

Treu, T. (2010). Strong Lensing by Galaxies. *ARA&A*, 48:87–125.

Treu, T., Brammer, G., Diego, J.M., Grillo, C., Kelly, P.L. et al (2016). “Refsdal” Meets Popper: Comparing Predictions of the Re-appearance of the Multiply Imaged Supernova Behind MACSJ1149.5+2223. *ApJ*, 817:60.

Treu, T. and Marshall, P.J. (2016). Time delay cosmography. *A&A Rev.*, 24:11.

Tsvetkova, V.S., Shulga, V.M. and Berdina, L.A. (2016). A simple method to determine time delays in the presence of microlensing: application to HE 0435-1112 and PG 1115+080. *MNRAS*, 461:3714–3723.

Tsvetkova, V.S., Vakulik, V.G., Shulga, V.M., Schild, R.E., Dudinov, V.N. et al (2010). PG1115+080: variations of the A2/A1 flux ratio and new values of the time delays. *MNRAS*, 406:2764–2776.

Tully, R.B. and Fisher, J.R. (1977). A new method of determining distances to galaxies. *A&A*, 54:661–673.

Tyson, J.A., Valdes, F. and Wenk, R.A. (1990). Detection of systematic gravitational lens galaxy image alignments - Mapping dark matter in galaxy clusters. *ApJ*, 349:L1–L4.

Udalski, A., Szymanski, M., Kaluzny, J., Kubiak, M. and Mateo, M. (1992). The Optical Gravitational Lensing Experiment. *Acta Astron.*, 42:253–284.

Unruh, S., Schneider, P. and Sluse, D. (2017). Ambiguities in gravitational lens models: the density field from the source position transformation. *A&A*, 601:A77.

Uttley, P., McHardy, I.M. and Vaughan, S. (2005). Non-linear X-ray variability in X-ray binaries and active galaxies. *MNRAS*, 359:345–362.

Vakulik, V.G., Shulga, V.M., Schild, R.E., Tsvetkova, V.S., Dudinov, V.N. et al (2009). Time delays in PG 1115+080: new estimates. *MNRAS*, 400:L90–L93.

van Engelen, A., Keisler, R., Zahn, O., Aird, K.A., Benson, B.A. et al (2012). A Measurement of Gravitational Lensing of the Microwave Background Using South Pole Telescope Data. *ApJ*, 756:142.

van Uitert, E., Joachimi, B., Joudaki, S., Heymans, C., Köhlinger, F. et al (2017). KiDS+GAMA: Cosmology constraints from a joint analysis of cosmic shear, galaxy-galaxy lensing and angular clustering. *ArXiv e-prints:1706.05004.*

Van Waerbeke, L., Mellier, Y., Erben, T., Cuillandre, J.C., Bernardeau, F. et al (2000). Detection of correlated galaxy ellipticities from CFHT data: first evidence for gravitational lensing by large-scale structures. *A&A*, 358:30–44.

Vanderriest, C., Schneider, J., Herpe, G., Chevreton, M., Moles, M. and Wlerick, G. (1989). The value of the time delay Delta t(A, B) for the 'double' quasar 0957+561 from optical photometric monitoring. *A&A*, 215:1–13.

Vuissoz, C., Courbin, F., Sluse, D., Meylan, G., Chantry, V. et al (2008). COSMOGRAIL: the COSmological MOnitoring of GRAvItational Lenses. VII. Time delays and the Hubble constant from WFI J2033-4723. *A&A*, 488:481–490.

Vuissoz, C., Courbin, F., Sluse, D., Meylan, G., Ibrahimov, M. et al (2007). COSMOGRAIL: the COSmological MOnitoring of GRAvItational Lenses. V. The time delay in SDSS J1650+4251. *A&A*, 464:845–851.

Walsh, D. (1989). *0957+561: The unpublished story*, pages 11–22. Springer Berlin Heidelberg, Berlin, Heidelberg.

Walsh, D., Carswell, R.F. and Weymann, R.J. (1979). 0957 + 561 A, B - Twin quasistellar objects or gravitational lens. *Nature*, 279:381–384.

Wambsganss, J. (1998). Gravitational Lensing in Astronomy. *Living Reviews in Relativity*, 1:12.

Wambsganss, J. (2006). Gravitational Microlensing. *ArXiv Astrophysics e-prints*.

Wambsganss, J., Bode, P. and Ostriker, J.P. (2005). Gravitational Lensing in a Concordance ΛCDM Universe: The Importance of Secondary Matter along the Line of Sight. *ApJ*, 635:L1–L4.

Wambsganss, J., Schmidt, R.W., Colley, W., Kundić, T. and Turner, E.L. (2000). Microlensing results from APO monitoring of the double quasar Q0957+561A,B between 1995 and 1998. *A&A*, 362:L37–L40.

Weinberg, D.H., Mortonson, M.J., Eisenstein, D.J., Hirata, C., Riess, A.G. and Rozo, E. (2013). Observational probes of cosmic acceleration. *Phys. Rep.*, 530:87–255.

Weymann, R.J., Latham, D., Roger, J., Angel, P., Green, R.F. et al (1980). The triple QSO PG1115+08 - Another probable gravitational lens. *Nature*, 285:641–643.

Wisotzki, L., Christlieb, N., Bade, N., Beckmann, V., Köhler, T., Vanelle, C. and Reimers, D. (2000). The Hamburg/ESO survey for bright QSOs. III. A large flux-limited sample of QSOs. *A&A*, 358:77–87.

Wisotzki, L., Koehler, T., Groote, D. and Reimers, D. (1996a). The Hamburg/ESO survey for bright QSOs. I. Survey design and candidate selection procedure. *A&AS*, 115:227.

Wisotzki, L., Koehler, T., Kayser, R. and Reimers, D. (1993). The new double QSO HE 1104-1805: Gravitational lens with microlensing or binary quasar? *A&A*, 278:L15–L18.

Wisotzki, L., Koehler, T., Lopez, S. and Reimers, D. (1996b). Discovery of a new gravitationally lensed QSO with broad absorption lines. *A&A*, 315:405–408.

Wisotzki, L., Schechter, P.L., Bradt, H.V., Heinmüller, J. and Reimers, D. (2002). HE 0435-1223: A wide separation quadruple QSO and gravitational lens. *A&A*, 395:17–23.

Wisotzki, L., Schechter, P.L., Chen, H.W., Richstone, D., Jahnke, K., Sánchez, S.F. and Reimers, D. (2004). HE 0047-1756: A new gravitationally lensed double QSO. *A&A*, 419:L31–L34.

Wisotzki, L., Wucknitz, O., Lopez, S. and Sorensen, A.N. (1998). First estimate of the time delay in HE 1104-1805. *A&A*, 339:L73–L76.

Wittman, D.M., Tyson, J.A., Kirkman, D., Dell'Antonio, I. and Bernstein, G. (2000). Detection of weak gravitational lensing distortions of distant galaxies by cosmic dark matter at large scales. *Nature*, 405:143–148.

Wong, K.C., Suyu, S.H., Auger, M.W., Bonvin, V., Courbin, F. et al (2017). H0LiCOW - IV. Lens mass model of HE 0435-1223 and blind measurement of its time-delay distance for cosmology. *MNRAS*, 465:4895–4913.

Wucknitz, O., Biggs, A.D. and Browne, I.W.A. (2004). Models for the lens and source of B0218+357: a LENSCLEAN approach to determine H_0. *MNRAS*, 349:14–30.

Xu, D., Springel, V., Sluse, D., Schneider, P., Sonnenfeld, A., Nelson, D., Vogelsberger, M. and Hernquist, L. (2016). The inner structure of early-type galaxies in the Illustris simulation. *ArXiv e-prints:1610.07605*.

Yoo, J., Kochanek, C.S., Falco, E.E. and McLeod, B.A. (2005). The Lens Galaxy in PG 1115+080 is an Ellipse. *ApJ*, 626:51–57.

York, T., Jackson, N., Browne, I.W.A., Wucknitz, O. and Skelton, J.E. (2005). The Hubble constant from the gravitational lens CLASS B0218+357 using the Advanced Camera for Surveys. *MNRAS*, 357:124–134.

Young, P., Gunn, J.E., Kristian, J., Oke, J.B. and Westphal, J.A. (1980). The double quasar Q0957 + 561 A, B - A gravitational lens image formed by a galaxy at Z = 0.39. *ApJ*, 241:507–520.

Zhang, B.R., Childress, M.J., Davis, T.M., Karpenka, N.V., Lidman, C., Schmidt, B.P. and Smith, M. (2017). A blinded determination of H_0 from low-redshift Type Ia supernovae, calibrated by Cepheid variables. *ArXiv e-prints:1706.07573*.

Zuntz, J., Paterno, M., Jennings, E., Rudd, D., Manzotti, A. et al (2015). CosmoSIS: Modular cosmological parameter estimation. *Astronomy and Computing*, 12:45–59.

Zwicky, F. (1933). Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6:110–127.

Zwicky, F. (1937a). Nebulae as gravitational lenses. *Phys. Rev.*, 51:290–290.

Zwicky, F. (1937b). On the probability of detecting nebulae which act as gravitational lenses. *Phys. Rev.*, 51:679–679.

Vivien BONVIN

Rue du Rocher 7
CH-2000 Neuchâtel
Tel:+41 79 630 63 68
Mail: vivien.bonvin@epfl.ch

Born July 30th 1989
27 years old
Swiss Nationality
Single

Work Experience

- **Ph.D** in Astrophysics at the Laboratory of Astrophysics of **EPFL**: 2013-2017
  - → Ph.D Thesis: *A competitive and complementary measurement of the Hubble Constant at 3.8% precision using Strong Gravitational Lensing.* From weekly telescope observations of lensed quasars carried over many years, I determined with high precision and accuracy new constraints on the expansion rate of the Universe using the Time Delay Strong Lensing technique.
  - → I authored two articles published in prestigious scientific journals about time-delay measurements from strongly lensed quasars monitoring and derived cosmological constraints. I am also the co-authors of six other publications.
- **International Physicists's Tournament (IPT)**: since 2012
  - → The IPT is a problem-based learning competition for university students that gather participants from all around the world. With other former Swiss participants, I organized the 5th and 6th editions of the IPT at EPFL that welcomed during one week a hundred of people from 11 different countries. I have been elected President of the IPT in 2014, ensuring the continuity in the organisation and growth of the tournament over the years
- **Teaching and Outreach**:
  - → I wrote, acted in and co-realized an outreach video on my research. The video was featured in a NASA / ESA joint press release on my work that made the headlines of several professional institutions and mainstream media around the world. 2017
  - → I am a member of the EPFL Massively Online Open Course (MOOC) *Introduction to Astrophysics* team, followed each semester by several thousands of people around the world. I am also teaching to bachelor and master degree students in Physics. since 2014

Education

- **Ecole Polytechnique Fédérale de Lausanne (EPFL)**: 2008-2013
  - → Bachelor & Master Degrees in Physics. I did my Master thesis in Astrophysics, using parallel computing and multiprocessing to perform high-resolution chemo-dynamical simulations of isolated spiral galaxies.
  - → I took part in 2011 as a bachelor student to the 3rd edition of the International Physicists' Tournament, in Moscow. I was a member of the first Swiss team that participated to this competition and we finished on the podium (3rd rank).
- **Lycée-Collège de la Planta**, Sion: 2003-2008
  - → I completed my Federal Maturity in Mathematics & Physics following a bilingual curriculum in French and German.

List of publications

- **Main author**:
  - → H0LiCOW V. New COSMOGRAIL time delays of HE0435-1223: H0 to 3.8% precision from strong lensing in a flat ΛCDM model. Published in *Monthly Notices of the Royal Astronomical Society*. 2017
  - → COSMOGRAIL XV. Assessing the achievability and precision of time-delay measurements. Published in *Astronomy & Astrophysics*. 2016
- **Co-author (selected publications)**:
  - → COSMOGRAIL XVI: Time delays for the quadruply imaged quasar DES J0408-5354 with high-cadence photometric monitoring. 2017
  - → Models of the strongly lensed quasar DES J0408-5354. 2017
  - → H0LiCOW I. H0 Lenses in COSMOGRAIL's Wellspring: Program Overview. 2017
  - → H0LiCOW III. Spectroscopic survey and galaxy-group identification of the strong gravitational lens system HE0435-1223. 2017
  - → H0LiCOW IV. Lens mass model of HE0435-1223 and blind measurement of its time-delay distance for cosmology. 2017
  - → Strong Lens Time Delay Challenge: II. Results of TDC1. 2015

See my full list of publications and associated statistics on the Astronomical Data System.

Technical & Computer Skills

- **Programming languages**:
  - → `Python`: very good knowledge, daily usage since 2012. Co-author of the time-serie analysis package `PyCS`, publicly available on github.
  - → `Django`: good knowledge. Author of the `IPT-connect` application, publicly available on github.
  - → `C++ / Matlab / HTML / PHP / CSS`: basic knowledge, self-taught.

- **Data analysis**:
  - → Image processing: in a data treatment pipeline I co-developed using advanced photometric measurement, PSF construction and image deconvolution techniques, I extracted a clean scientific signal from raw images taken by multiple telescopes around the world.
  - → Time serie anaalysis: using advanced curve fitting techniques such as BOK splines and Gaussian process regressions, I analysed time series of irregular sampling and reproduced simulated mock data sets mimicking the behavior of the original data to assess the robustness of my measurements.
  - → Statistical analysis: with the help of robust statistical tools such as Monte Carlo Markov Chain processes applied in a Bayesian framework or Frequentist inference and probability theory, I derived meaningful and robust constraints from my data sets.

Languages

English	**French**	**German**	**Others**
Fluent. High-school knowledge plus one month of specific language course in Cambridge, UK (2012). Used daily (spoken and written) since 2013.	Mother tongue	High-school knowledge	Spanish, Polish: basic knowledge through self-teaching and courses in language schools